Towards segment-based recognition of argumentation structure in short texts

Andreas Peldszus
Supervisor: Manfred Stede

Applied Computational Linguistics, University of Potsdam

1st ACL WS on Argumentation Mining, June 26, 2014
What makes argumentation mining so hard?

- lots of text available, but only few arguments
- argumentative strategies vary across texts genres, topic, author
- understanding inferences may require very topic-specific background knowledge
- implicitness of argumentation
  - suppressed premisses
  - linguistic markedness
  - rhetorically gimmicks
Data: pro & contra commentaries

Source:
- pro & contra newspaper commentaries
- in Potsdam Commentary Corpus
  [Stede, 2004] [Stede and Neumann, 2014]

Properties:
+ lots of arguments
+ rather explicitly marked argumentation
  - special background knowledge required
  - main claim may be implicit
  - full range of persuasive ‘tricks’ professional writers have to offer
Data: microtexts

Source:

- 23 texts: hand-crafted, covering different arg. configurations
- 92 texts: collected in a controlled text generation experiment

Properties:

+ each segment is arg. relevant
+ explicit main claim
+ at least one possible objection considered

A (translated) example

[ Energy-saving light bulbs contain a considerable amount of toxic substances. ]₁ [ A customary lamp can for instance contain up to five milligrams of quicksilver. ]₂ [ For this reason, they should be taken off the market, ]₃ [ unless they are virtually unbreakable. ]₄ [ This, however, is simply not case. ]₅
Outline

1. Dataset Generation
2. Scheme
3. Annotation Study
4. Automatic Recognition
Outline

1. Dataset Generation
2. Scheme
3. Annotation Study
4. Automatic Recognition
Generation of argumentative micro-texts: Collecting

Text generation experiment:
- 23 probands (of varying age, education and occupation)
- discuss a controversial issue (recent political, moral, everyday’s life questions) in a short text
- max. 5 texts per proband

Requirements:
- length of five segments
- all segments argumentatively relevant
- at least one possible objection to be considered
- text understandable for readers without knowing the issue question
Generation of argumentative micro-texts: Collecting

Text generation experiment:

- 23 probands (of varying age, education and occupation)
- discuss a controversial issue (recent political, moral, everyday’s life questions) in a short text
- max. 5 texts per proband

Requirements:

- length of five segments
- all segments argumentatively relevant
- at least one possible objection to be considered
- text understandable for readers without knowing the issue question
Generation of argumentative micro-texts: Dataset

Resulting Dataset:

- 100 authentic texts
- 92 after cleanup
- plus 23 artificial texts

= 115 texts, 579 segments, now annotated with argumentation graphs!
Outline

1 Dataset Generation
2 Scheme
3 Annotation Study
4 Automatic Recognition
Scheme: A theory of argumentation structure

Freeman’s theory, revised & slightly generalized:

- 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)
Freeman’s theory, revised & slightly generalized:

- 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)

Further complete annotation of authentic text:
- glue(3,4) – unitizing ADUs from EDUs
- skip(10) – arg. irrelevant segments
- join(5,13) – restatements
Outline

1 Dataset Generation
2 Scheme
3 Annotation Study
4 Automatic Recognition
Annotation study

expert annotators: guideline authors + postdoc + student

[This study]
Annotation study

naive, min. trained annotators: 26 undergrad students

expert annotators: guideline authors + postdoc + student

[Peldszus and Stede, 2013a]

[This study]
Outline

1 Dataset Generation
2 Scheme
3 Annotation Study
4 Automatic Recognition
Modelling micro-texts: Segment-wise classification

<table>
<thead>
<tr>
<th>prediction</th>
<th>text segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:</td>
<td>Energy-saving light bulbs ...</td>
</tr>
<tr>
<td>2:</td>
<td>A customary lamp can for ins...</td>
</tr>
<tr>
<td>3:</td>
<td>For this reason, they should ...</td>
</tr>
<tr>
<td>4:</td>
<td>unless they are virtually unbr...</td>
</tr>
<tr>
<td>5:</td>
<td>This, however, is simply not ...</td>
</tr>
</tbody>
</table>

Simple, supervised machine-learning approach, inspired by Argumentative Zoning models.
Modelling micro-texts: Segment-wise classification

Simple, supervised machine-learning approach, inspired by Argumentative Zoning models.

PSN(n+2)

1: Energy-saving light bulbs ...
2: A customary lamp can for ins...
3: For this reason, they should ...
4: unless they are virtually unbr...
5: This, however, is simply not ...

prediction text segment
### Modelling micro-texts: Segment-wise classification

<table>
<thead>
<tr>
<th>prediction</th>
<th>text segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSN(n+2)</td>
<td>1: Energy-saving light bulbs ...</td>
</tr>
<tr>
<td>PSE(n-1)</td>
<td>2: A customary lamp can for ...</td>
</tr>
<tr>
<td></td>
<td>3: For this reason, they should ...</td>
</tr>
<tr>
<td></td>
<td>4: unless they are virtually unbr...</td>
</tr>
<tr>
<td></td>
<td>5: This, however, is simply not ...</td>
</tr>
</tbody>
</table>

Simple, supervised machine-learning approach, inspired by Argumentative Zoning models.
Modelling micro-texts: Segment-wise classification

prediction | text segment
--- | ---
PSN(n+2) | 1: Energy-saving light bulbs …
PSE(n-1) | 2: A customary lamp can for ins…
PT(0) | 3: For this reason, they shou…
 | 4: unless they are virtually unbr…
 | 5: This, however, is simply not …

Simple, supervised machine-learning approach, inspired by Argumentative Zoning models.
Modelling micro-texts: Segment-wise classification

<table>
<thead>
<tr>
<th>prediction</th>
<th>text segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSN(n+2)</td>
<td>1: Energy-saving light bulbs ...</td>
</tr>
<tr>
<td>PSE(n-1)</td>
<td>2: A customary lamp can for ins...</td>
</tr>
<tr>
<td>PT(0)</td>
<td>3: For this reason, they should ...</td>
</tr>
<tr>
<td>OAU(r-3)</td>
<td>4: unless they are virtually u...</td>
</tr>
<tr>
<td></td>
<td>5: This, however, is simply not ...</td>
</tr>
</tbody>
</table>

Simple, supervised machine-learning approach, inspired by Argumentative Zoning models.
Modelling micro-texts: Segment-wise classification

<table>
<thead>
<tr>
<th>prediction</th>
<th>text segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSN(n+2)</td>
<td>1: Energy-saving light bulbs ...</td>
</tr>
<tr>
<td>PSE(n-1)</td>
<td>2: A customary lamp can for ins...</td>
</tr>
<tr>
<td>PT(0)</td>
<td>3: For this reason, they should ...</td>
</tr>
<tr>
<td>OAU(r-3)</td>
<td>4: unless they are virtually unbr...</td>
</tr>
<tr>
<td>PAR(n-1)</td>
<td>5: This, however, is simply n...</td>
</tr>
</tbody>
</table>

Simple, supervised machine-learning approach, inspired by Argumentative Zoning models.
Modelling micro-texts: Segment-wise classification

Simple, supervised machine-learning approach, inspired by Argumentative Zoning models.

<table>
<thead>
<tr>
<th>prediction</th>
<th>text segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSN(n+2)</td>
<td>1: Energy-saving light bulbs …</td>
</tr>
<tr>
<td>PSE(n-1)</td>
<td>2: A customary lamp can for ins…</td>
</tr>
<tr>
<td>PT(0)</td>
<td>3: For this reason, they should …</td>
</tr>
<tr>
<td>OAU(r-3)</td>
<td>4: unless they are virtually unbr…</td>
</tr>
<tr>
<td>PAR(n-1)</td>
<td>5: This, however, is simply not …</td>
</tr>
</tbody>
</table>

Simple, supervised machine-learning approach, inspired by Argumentative Zoning models.
Modelling micro-texts: Features

- Lemma unigrams (with ±1 window)
- Lemma bigrams
- First three lemma
- Part of speech tags (with ±1 window)
- Main verb morphology, e.g. mood & tempus
- Dependency syntax triples, lemma-based
- Dependency syntax triples, POS-based
- Discourse markers and marked relations from DimLex [Stede, 2002] (with ±1 window)
- Negation marker presence [Warzecha, 2013]
- Sentiment, sum of all pos. and neg. values, according to SentiWS [Remus et al., 2010]
- Segment position in text (relative)
Modelling micro-texts: Models

Baselines:

- **Majority**: Weka’s [Hall et al., 2009] ZeroR
- **One Rule**: Weka’s OneR with standard parameters

Simple Models:

- **Naïve Bayes**: Weka’s Naïve Bayes, features with information gain \( \neq 0 \) are excluded
- **SVM**: Weka’s wrapper to LibLinear [Fan et al., 2008] with the Crammer and Singer SVM type and standard wrapper parameters
- **MaxEnt**: MaxEnt toolkit [Zhang, 2004], 50 iterations, L-BFGS, no Gaussian prior
- **CRF**: Mallet [McCallum, 2002]. SimpleTagger interface with standard parameters
Modelling micro-texts: Results ($F_1$)
Modelling micro-texts: Results ($F_1$)

Towards segment-based recognition of arg. structure

Andreas Peldszus (Uni Potsdam)
Modelling micro-texts: Results ($F_1$)

![Graph](image)

- **target**
- **comb**
- **type**
- **typegen**
- **role**

- **Majority**
- **OneRule**
- **Best**

Andreas Peldszus (Uni Potsdam)
Towards segment-based recognition of arg. structure
ArgMining 2014 17 / 27
Modelling micro-texts: Results ($F_1$)

- role + type + comb + target
- role + type
- role + typegen
- target
- comb
- type
- typegen
- role

<table>
<thead>
<tr>
<th></th>
<th>Majority</th>
<th>OneRule</th>
<th>Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>role + typegen</td>
<td>80</td>
<td>70</td>
<td>60</td>
</tr>
<tr>
<td>role + type</td>
<td>70</td>
<td>60</td>
<td>50</td>
</tr>
<tr>
<td>role + type</td>
<td>60</td>
<td>50</td>
<td>40</td>
</tr>
<tr>
<td>role + type + comb + target</td>
<td>50</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td>role + type + comb + target</td>
<td>40</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>role + type + comb + target</td>
<td>30</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>role + type + comb + target</td>
<td>20</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

Towards segment-based recognition of arg. structure

Andreas Peldszus (Uni Potsdam)
Modelling micro-texts: Results ($F_1$)

- role+type+comb+target
- role+type
- role+typegen
- target
- comb
- type
- typegen
- role

Legend:
- Majority
- OneRule
- Best
Modelling micro-texts: Results ($\kappa$)

Towards segment-based recognition of arg. structure
Modelling micro-texts: Results - class wise

<table>
<thead>
<tr>
<th>label</th>
<th>precision</th>
<th>recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT</td>
<td>75±12</td>
<td>74±13</td>
<td>74±11</td>
</tr>
<tr>
<td>PSN</td>
<td>65±8</td>
<td>79±7</td>
<td>71±6</td>
</tr>
<tr>
<td>PSE</td>
<td>1±6</td>
<td>1±6</td>
<td>1±6</td>
</tr>
<tr>
<td>PAR</td>
<td>12±29</td>
<td>12±27</td>
<td>11±24</td>
</tr>
<tr>
<td>PAU</td>
<td>57±26</td>
<td>49±24</td>
<td>50±22</td>
</tr>
<tr>
<td>OSN</td>
<td>1±12</td>
<td>1±12</td>
<td>1±12</td>
</tr>
<tr>
<td>OAR</td>
<td>54±18</td>
<td>42±16</td>
<td>46±13</td>
</tr>
<tr>
<td>OAU</td>
<td>8±27</td>
<td>7±23</td>
<td>7±23</td>
</tr>
</tbody>
</table>

MaxEnt class-wise results on the ‘role+type’ level: Percent average and standard deviation in 10 repetitions of 10-fold cross-validation of Precision, Recall and F1-score.
## Modelling micro-texts: Results - class wise

<table>
<thead>
<tr>
<th>label</th>
<th>precision</th>
<th>recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT</td>
<td>75±12</td>
<td>74±13</td>
<td>74±11</td>
</tr>
<tr>
<td>PSN</td>
<td>65±8</td>
<td>79±7</td>
<td>71±6</td>
</tr>
<tr>
<td>PSE</td>
<td>1±6</td>
<td>1±6</td>
<td>1±6</td>
</tr>
<tr>
<td>PAR</td>
<td>12±29</td>
<td>12±27</td>
<td>11±24</td>
</tr>
<tr>
<td>PAU</td>
<td>57±26</td>
<td>49±24</td>
<td>50±22</td>
</tr>
<tr>
<td>OSN</td>
<td>1±12</td>
<td>1±12</td>
<td>1±12</td>
</tr>
<tr>
<td>OAR</td>
<td>54±18</td>
<td>42±16</td>
<td>46±13</td>
</tr>
<tr>
<td>OAU</td>
<td>8±27</td>
<td>7±23</td>
<td>7±23</td>
</tr>
</tbody>
</table>

MaxEnt class-wise results on the ‘role+type’ level: Percent average and standard deviation in 10 repetitions of 10-fold cross-validation of Precision, Recall and F1-score.
Modelling micro-texts: Results - feature ablation

<table>
<thead>
<tr>
<th>Features</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F_1$</td>
</tr>
<tr>
<td>all</td>
<td>60±5</td>
</tr>
<tr>
<td>only lemma unigrams</td>
<td>55±5</td>
</tr>
<tr>
<td>only lemma bigrams</td>
<td>53±5</td>
</tr>
<tr>
<td>only discourse markers</td>
<td>53±6</td>
</tr>
<tr>
<td>only first three lemma</td>
<td>52±6</td>
</tr>
<tr>
<td>only dependencies lemma</td>
<td>47±4</td>
</tr>
<tr>
<td>only pos</td>
<td>45±6</td>
</tr>
<tr>
<td>only dependencies pos</td>
<td>41±6</td>
</tr>
<tr>
<td>only main verb morph</td>
<td>39±4</td>
</tr>
<tr>
<td>only segment position</td>
<td>25±10</td>
</tr>
<tr>
<td>only negation marker</td>
<td>19±8</td>
</tr>
<tr>
<td>only sentiment</td>
<td>15±11</td>
</tr>
</tbody>
</table>

Feature ablation tests on the role+type level: Percent average and standard deviation in 10 repetitions of 10-fold cross-validation of micro averages of F1-scores, and Cohen’s $\kappa$. 
Outlook & Future Work

Data:
- generate more microtexts (crows-sourced text generation on the way)
- annotate more newspaper texts (new tool: GraPAT) [Sonntag and Stede, 2014]

Models:
- rerank predicted labels by graph-validity constraints (done)
- separate, individually tuned classifiers for different graph aspects
- ...

Features:
- automatic disambiguation of discourse markers
- use semantic similarity, contrastive word pairs for less-marked transitions
Outlook & Future Work

Data:

• generate more microtexts (crow-sourced text generation on the way)
• annotate more newspaper texts (new tool: GraPAT) [Sonntag and Stede, 2014]

Models:

• rerank predicted labels by graph-validity constraints (done)
• separate, individually tuned classifiers for different graph aspects
• …

Features:

• automatic disambiguation of discourse markers
• use semantic similarity, contrastive word pairs for less-marked transitions
Outlook & Future Work

Data:
- generate more microtexts (crowd-sourced text generation on the way)
- annotate more newspaper texts (new tool: GraPAT) [Sonntag and Stede, 2014]

Models:
- rerank predicted labels by graph-validity constraints (done)
- separate, individually tuned classifiers for different graph aspects
- ...

Features:
- automatic disambiguation of discourse markers
- use semantic similarity, contrastive word pairs for less-marked transitions
Thank You!


Annotation study

Rewrite graphs as a list of (relational) segment labels

<table>
<thead>
<tr>
<th>role</th>
<th>P(proponent)</th>
<th>Oponent</th>
</tr>
</thead>
<tbody>
<tr>
<td>typegen</td>
<td>T(thesis)</td>
<td>S(support)</td>
</tr>
<tr>
<td>type</td>
<td>N(ormal)</td>
<td>E(example)</td>
</tr>
<tr>
<td>combined</td>
<td>S(simple)</td>
<td>C(combined)</td>
</tr>
</tbody>
</table>

1: PSNS(3)
2: PSES(1)
3: PT()
4: OARS(3)
5: PARS(4)
### Modelling micro-texts: Annotated corpus

<table>
<thead>
<tr>
<th>level</th>
<th>role</th>
<th>typegen</th>
<th>type</th>
<th>comb</th>
<th>target</th>
<th>role+type</th>
</tr>
</thead>
<tbody>
<tr>
<td>labels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P (454)</td>
<td>T (115)</td>
<td>T (115)</td>
<td>/ (115)</td>
<td>n-4 (26)</td>
<td>PT (115)</td>
</tr>
<tr>
<td></td>
<td>O (125)</td>
<td>S (286)</td>
<td>SN (277)</td>
<td>S (426)</td>
<td>n-3 (52)</td>
<td>PSN (265)</td>
</tr>
<tr>
<td></td>
<td>A (178)</td>
<td>SE (9)</td>
<td>SE (9)</td>
<td>C (38)</td>
<td>n-2 (58)</td>
<td>PSE (9)</td>
</tr>
<tr>
<td></td>
<td>AU (66)</td>
<td>AR (112)</td>
<td>AR (112)</td>
<td></td>
<td>n-1 (137)</td>
<td>PAR (12)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0 (115)</td>
<td>PAU (53)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>n+1 (53)</td>
<td>OSN (12)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>n+2 (35)</td>
<td>OSE (0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>r-1 (54)</td>
<td>OAR (100)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>r-2 (7)</td>
<td>OAU (13)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

# of lbls 2 3 5 3 16 9

Label distribution on the basic levels and for illustration on the complex ‘role+type’ level.
Generation of argumentative micro-texts: Topics

Top 5 chosen topics:

- Should the fine for leaving dog excrements on sidewalks be increased?
- Should shopping malls generally be allowed to open on Sundays?
- Should Germany introduce the death penalty?
- Should public health insurance cover treatments in complementary and alternative medicine?
- Should only those viewers pay a TV licence fee who actually want to watch programs offered by public broadcasters?