DISCOURSE

From Modeling of Discourse to Modeling with Discourse and the First Visit to Argument Mining

CS 3730 Class discussion
2/13/2014

Huy Nguyen
Discourse analysis

• “not only study language use 'beyond the sentence boundary', but also prefer to analyze 'naturally occurring' language use, and not invented examples” [Wikipedia]

• Not only linguistics
  • Applied linguistics, pragmatics, rhetoric

• But also sociolinguistics
  • Interactional sociolinguistics, code-switching, variation analysis
  • From the course I took: how your language use reveals your identity
Today’s discussion

• Modeling of Discourse
  • A Weakly-supervised Approach to AZ (2011)

• Modeling with Discourse
  • Predicting the Presence of Discourse Connectives (2013)
  • Exploiting DA for Article-Wide Temporal Classification (2013)

• Argument Mining
  • Classifying Arguments by Scheme (2011)

• Link them all
A WEAKLY-SUPERVISED APPROACH TO ARGUMENTATIVE ZONING OF SCIENTIFIC DOCUMENTS

Guo et al. EMNLP 2011
Classified

• Measurement paper
  • Following work of Guo et al. (2010)
  • Weakly supervised to address limited labeled data, model porting

• Data domain
  • Guo et al.
  • Medline abstracts annotated according to AZ

• Contribution
  
  Our experiments show that weakly-supervised learning can be used to identify AZ in scientific documents with good accuracy when only a limited amount of labeled data is available. This is helpful thinking of the real-world application and porting of the approach to different tasks and domains. To the best of our knowledge, no previous work has been done on weakly-supervised learning of information structure according to schemes of the type we have focused on (Teufel and Moens, 2002; Mizuta et al., 2006; Lin et al., 2006; Hirohata et al., 2008; Shatkay et al., 2008; Liakata et al., 2010).
Which is more interesting

- The task
  - Identifying the Information Structure of Scientific Abstracts
  - According to AZ
- The techniques
  - Active SVM (2), Transductive SVM, Semi-supervised CRF
  - Future work
- The task is what matters and earns them publications
PREDICTING THE PRESENCE OF DISCOURSE CONNECTIVES

Patterson and Kehler EMNLP 2013
Classified

- Idea paper
  - Predicts the presence or omission of a lexical connective between two clauses
  - To yield fluid transitions between clauses

- Data domain
  - Penn Discourse Treebank

- Application domain
  - Natural language generation, summarization
**PDTB (Super brief snapshot)**

- Connectives and arguments
  - Explicit vs. implicit

- Senses
  - Organized hierarchically
  - 4-tag top level: temporal, contingency, comparison, expansion
  - Up to 3-level depth

(7) *Use of dispersants was approved when a test on the third day showed some positive results, officials said.* (CONTINGENCY:Cause:reason) (1347)

(123) A Lorillard spokeswoman said, “*This is an old story. Implicit = IN FACT We’re talking about years ago before anyone heard of asbestos having any questionable properties.*” (EXPANSION:Restatement:specification) (0003)
Difficulty of the problem

- Presence or omission of connectives
  - Binary classification

- More challenging problems

- Error analysis

multi-class classification [Fan]

just [considered] two adjacent arguments [Zahra]
detect the relation type between two clauses [Changsheng]

interesting [Fataneh, Zahra]
cases where connectives are optional [Phuong]
Feasibility of the features

• Relation-level
  • Connect, AttMismatch, Financial

• Argument-level
  • Supplementary, length, #clauses, content-word ratio, pronoun ratio, FirstA2Pron

• Discourse-level
  • Prior semantic type, relation dependency (shared/embedded),

*How [Connect] was constructed? [Phuong]*

*know the relation between two arguments and also the correct connectives [Zahra]*
EXPLOITING DISCOURSE ANALYSIS FOR ARTICLE-WIDE TEMPORAL CLASSIFICATION

Ng et al. EMNLP 2013
Classified

• Idea paper
  • Temporal relations between pairs of events beyond the sentence level
  • Exploit discourse structure and topic segmentation

• Data domain
  • Newswire articles from ACE 2005 corpus (cf. Do et al. 2012)

• Application domain
  • Information extraction and text summarization, e.g. event timeline
RST discourse

- Originally developed as part of studies of computer-based text generation
- RST offers an explanation of the coherence of texts
  - An evident role for every part
- Nucleus::Satellite Relations
- Multinuclear Relations
RST v. PDTB

learned a lot about discourse [Huma]

RESULT

Max switched off the light.

The room was pitch dark.

CONTINGENCY :: CAUSE

arg1
Max switched off the light.

arg2
The room was pitch dark.
Making use of discourse

- RST discourse framework (by Feng and Hirst 2012)
- PDTB discourse relation (by Lin et al. 2013)
  - Path of relations connecting two events

*transitivity property of temporal relationships* [Changsheng]

Problem [of] the sparsity of temporal relations [Huichao]
Making use of discourse (cont.)

- Topical text segmentation (by Kazantseva and Szpakowicz 2011)
  - Index of segments

motivation behind using text segmentation [Zahra]

not that intuitive but effective [Fan]
CLASSIFYING ARGUMENTS BY SCHEME

Feng and Hirst ACL HLT 2011
Classified

• Idea/Measurement paper
  • Identify argumentation schemes to reconstruct the enthymemes in an argument (unstated/implicitly stated premises)

• Data domain
  • Araucaria dataset: annotated for conclusion, premises, scheme

• Application domain
  • Argument understanding, argument supporting/attacking
Sample annotated text in Araucaria

If we stop the free creation of art, we will stop the free viewing of art.

<table>
<thead>
<tr>
<th>PROP identifier=&quot;C&quot; missing=&quot;yes&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROPTEXT offset=&quot;-1&quot;&gt; The prohibition of the free creation of art should not be brought about. &lt;/PROPTEXT&gt;</td>
</tr>
</tbody>
</table>
| INSCHHEME scheme="Argument from Consequences" schid="0" />

<table>
<thead>
<tr>
<th>PROP identifier=&quot;A&quot; missing=&quot;no&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROPTEXT offset=&quot;0&quot;&gt; If we stop the free creation of art, we will stop the free viewing of art. &lt;/PROPTEXT&gt;</td>
</tr>
</tbody>
</table>
| INSCHHEME scheme="Argument from Consequences" schid="0" />

<table>
<thead>
<tr>
<th>PROP identifier=&quot;B&quot; missing=&quot;yes&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROPTEXT offset=&quot;-1&quot;&gt; The prohibition of free viewing of art is not acceptable. &lt;/PROPTEXT&gt;</td>
</tr>
</tbody>
</table>
| INSCHHEME scheme="Argument from Consequences" schid="0" />

<table>
<thead>
<tr>
<th>LA</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>AU</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>AU</td>
</tr>
</tbody>
</table>
The three problems

• Premise/conclusion classifier
• Scheme classifier
• Infer enthymeme

Not all schemes [Fataneh, Wecan]

Contracy type feature can be derived when all the premises are present [Huichao]
Discourse features

- Location-based features of conclusion/premises
- Type of argumentation structure: linked/convergent
- Cue words/phrases of different argument schemes

*using words and/or rules would work well [Phuong]*
CONCLUSION
Link them all

• What we think of DA
  • Role of text blocks
    • Overall text for global roles (Guo et al. 2010, 2011)

• Relation between text blocks
  • Adjacent segments for local relations
  • Semantic types (Feng and Hirst 2011)
  • Cue words/phrase (Madnani et al. 2012)

• Layering discourse frameworks
  • Argumentative discourse (Feng and Hirst 2011)
  • Scientific discourse (Guo et al. 2010, 2011)
Link them all (cont.)

- Where we see DA
  - Fluid transition in NLG (Patterson and Kehler 2013)
  - Temporal relation in article-wide (Ng et al. 2013)
  - Argument schemes (Feng and Hirst 2011)

- How we do DA
  - Pure likes RST, PDTB
    - Supervised, weakly supervised, rule-based
    - Lexical, syntactic, statistic features
    - In future: advanced ML techniques, sophisticate feature

- Layered likes Argumentation schemes, Scientific discourse
Think of your research

How DA benefits your research? [Huy, Phuong]
Thank you!

- For all of your comments, participation