Combining Intra- and Multi-sentential Rhetorical Parsing for Document-level Discourse Analysis
Shafiq Joty, Giuseppe Carenini, Raymond T. Ng and Yashar Mehdad

The Problem

[But he added], "some people use the purchasers' index as a leading indicator."

[Some use it as a coincident indicator. But the thing it's supposed to measure, ...

Discourse structures may violate sentence boundaries.

5% and 12% sentences

HILDA

TSP SW outperforms TSP 1

Chain

Cross

Allows us to balance the data.

In Sliding

ILP

Into EDUs

Text

1. Parsing model

assigns probabilities to all possible DTs

R ranges over set of relations

1(a). Intra-sentential Parsing model

[Joty et al, EMNLP’12]

Relation at level $i$

$R \in \{1 \ldots M\}$

Structure at level $i$

$S \in \{0, 1\}$

Spans at level $i$

Obtaining probabilities: Apply the DCRF model to all possible sequences at different levels and compute posterior marginals of relation-structure pairs.

1(b). Multi-sentential Parsing model

- Chain-structured intra-sentential model doesn’t scale up.
- Break the chain structure.
- Allows us to balance the data.

Obtaining probabilities: Apply the CRF model to all possible adjacent units at different levels and compute posterior marginals of relation-structure pairs.

Combining Intra- and Multi-sentential

1S-1S (1 Sentence- 1 Sub-tree)

The intra-sentential parser first constructs a DT for every sentence, then the multi-sentential parser builds full DT on top of those.

Sliding Window

- Discourse structures may violate sentence boundaries.
- 5% and 12% sentences don’t have a DT in RST-DT and Instructional corpora, respectively.
- Often the units are connected to adjacent sentences.
- In Sliding Window, the intra-sentential parser builds a DT for each window of two consecutive sentences and then consolidates the decisions.

Consolidation: 3 cases

a) Same in both
b) Different but no cross
c) Cross

Experiments

Experimental Setup

Corpora
- RST-DT: 347 train, 38 test; 18 relations (41 with Nucleus Satellite attached).
- Instructional: 151 train, 25 test; 26 relations (76 with Nuc. Sat. attached).

Systems Compared with
- HILDA (Hernault et al., 2010) on RST-DT corpus.

Parsing Results

Used standard unlabeled (span) and labeled (nuclearity, relation) metrics [Marcu’00]

<table>
<thead>
<tr>
<th>RST-DT</th>
<th>Instructional</th>
</tr>
</thead>
<tbody>
<tr>
<td>RST-DT</td>
<td></td>
</tr>
<tr>
<td>HILDA</td>
<td>74.7</td>
</tr>
<tr>
<td>TSP 1-1</td>
<td>82.5</td>
</tr>
<tr>
<td>TSP SW</td>
<td>82.7</td>
</tr>
<tr>
<td>Human</td>
<td>86.7</td>
</tr>
<tr>
<td>ILP</td>
<td>70.35</td>
</tr>
<tr>
<td>TSP 1-1</td>
<td>79.67</td>
</tr>
<tr>
<td>TSP SW</td>
<td>80.88</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Span</th>
<th>Human</th>
<th>ILP</th>
<th>TSP 1-1</th>
<th>TSP SW</th>
</tr>
</thead>
<tbody>
<tr>
<td>44.3</td>
<td>55.7</td>
<td>55.7</td>
<td>65.8</td>
<td>35.44</td>
</tr>
<tr>
<td>35.44</td>
<td>43.58</td>
<td>43.58</td>
<td>43.58</td>
<td>43.58</td>
</tr>
</tbody>
</table>

- Our parsers TSP 1-1 and TSP SW outperform existing systems by a wide margin.
- TSP SW outperforms TSP 1-1 only in span (i.e., tree structure).
- TSP SW tends to induce noise from its neighbors for relation labels.

Reference: