1. Rhetorical Structure Theory[1]

- A discourse tree representation of the full text.
  - Leaves: elementary discourse units (EDUs).
  - Internal nodes: concatenation of continuous EDUs, with discourse relations among them labeled.
- Example
  
  [Catching up with commercial competitors in retail banking and financial services,]_{e_1} [they argue,]_{e_2} [will be difficult,]_{e_3} [particularly if market conditions turn sour.]_{e_4}

2. Two-stage Discourse Parsing

- Joint modeling of the structure and the relation of adjacent text units.
- CKY-like parsing algorithm to build the discourse tree from bottom up.

**Pros**
- Takes into account the interaction between structures and relations.
- Globally optimal tree.

**Cons**
- Inefficient for large documents. CKY-parsing: $O(n^3)$ time complexity.

3. Overall Workflow

4. Linear-chain CRFs with Constraints as Local Models

- Two-stage pipeline in local models:
  - Decompose $P(\text{structure}, \text{relation})$ into $P(\text{structure}) \times P(\text{relation} | \text{structure})$.
  - First identify the pair of adjacent text units to be related, then assign the relation to the pair.
  - Efficient in practice: $O(n)$ time complexity.

5. Experiments

- Features
  - Organization, textual, n-gram, dominance, contextual, substructure, entity transition, cue phrases, and post-editing features.
- Data
  - The RST Discourse Treebank: 347 for training and 38 for testing.
  - # of sentences: 2 to 187, average 26.
- Evaluation
  - Parsing accuracy using constituent precision and recall[1].
  - Parsing time.

**References**


6. Results

- Parsing accuracy

<table>
<thead>
<tr>
<th>Model</th>
<th>Span</th>
<th>Nuclearity</th>
<th>Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joty et al., 2013[2]</td>
<td>82.5</td>
<td>68.4</td>
<td>55.7</td>
</tr>
<tr>
<td>Feng and Hirst, 2012[4]</td>
<td>82.8</td>
<td>67.1</td>
<td>52.0</td>
</tr>
<tr>
<td>Ours (no post-editing)</td>
<td>84.9*</td>
<td>69.9*</td>
<td>57.2*</td>
</tr>
<tr>
<td>Ours</td>
<td>85.7**</td>
<td>71.0**</td>
<td>58.2**</td>
</tr>
<tr>
<td>Human</td>
<td>88.7</td>
<td>77.7</td>
<td>65.8</td>
</tr>
</tbody>
</table>

* Significantly better than Joty et al., 2013 (p < .01).
** Significantly better than ours (no post-editing) (p < .01).

- Parsing time in seconds

<table>
<thead>
<tr>
<th>Model</th>
<th>Min</th>
<th>Max</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feng and Hirst, 2012[4]</td>
<td>0.42</td>
<td>124.86</td>
<td>11.19</td>
</tr>
<tr>
<td>Ours (no post-editing)</td>
<td>0.05</td>
<td>40.57</td>
<td>5.52</td>
</tr>
<tr>
<td>Ours</td>
<td>0.12</td>
<td>84.72</td>
<td>10.71</td>
</tr>
</tbody>
</table>

Parsing time excludes the time for necessary pre-processing.

Acknowledgements: We thank Professor Gerald Penn and the reviewers for their valuable advice and comments. This work was financially supported by the Natural Sciences and Engineering Research Council of Canada and by the University of Toronto.