A Joint Phrasal and Dependency Model for Paraphrase Alignment

Kapil Thadani, Scott Martin and Michael White
kapil@cs.columbia.edu, {scott,mwhite}@ling.ohio-state.com

Monolingual Alignment

- **Goal**: identify semantically equivalent words and phrases across pairs of text segments
- Monolingual variant of MT alignment
- Useful for paraphrasing, entailment, sentence fusion, QA matching, plagiarism detection etc

Approach

- Supervised structured prediction for pairwise text alignment
- Joint inference to simultaneously align phrases and dependency arcs
- Assumes the score of an alignment factors into scores for phrase edits and arc edits

Joint inference via ILP

Indicator variables for phrase edits \( y \), arc edits \( z \) and token pair alignments \( x \)

Objective: \[
\max \sum y^T_w \Phi(y) + \sum z^T_w \Phi(z)
\]
subject to constraints:
- Only one active \( y \) per token, one active \( z \) per dependency
- Each active \( x \) participates in exactly one active \( y \)
- Each \( z \) activates two \( x \) to pair off its heads and dependents respectively and, conversely, is activated if these \( x \) are active

Challenges

- Phrase-based representations are natural but paraphrase recall is problematic
- ~65% of token alignments in Edinburgh training corpus supported by dependencies

Edinburgh Paraphrase Corpus

- Human-aligned paraphrase corpus (Cohn et al., 2008) from three sources:
  1. Multiple Translation Chinese corpus
  2. Multiple Jules Verne translations
  3. MSR paraphrase corpus
- Retokenized, truecased, NEs collapsed
- 715 training + 305 test instances (70:30 split)

Baseline 1. Meteor (Denkowski & Lavie, 2011)
- MT evaluation metric
- best configuration over training set (**max-accuracy**) Results (F₁)

<table>
<thead>
<tr>
<th></th>
<th>Tokens</th>
<th>Phrase-based</th>
<th>Phrase+Arc</th>
</tr>
</thead>
<tbody>
<tr>
<td>SURE</td>
<td>75.49%</td>
<td>77.85%</td>
<td>79.20%</td>
</tr>
<tr>
<td>SURE + POSSIBLE</td>
<td>73.22%</td>
<td>75.37%</td>
<td>77.57%</td>
</tr>
</tbody>
</table>

Baseline 2. Phrase-based MANLI (Thadani & McKeown, 2011)
- Supervised aligner with ILP inference
- Improvement on MANLI (MacCartney et al., 2008) which outperforms GIZA++ (Och & Ney 2003), HMMs (Liang et al., 2006) and Stanford’s RTE aligner (Chambers et al., 2007)

<table>
<thead>
<tr>
<th></th>
<th>Tokens</th>
<th>Phrase-based</th>
<th>Phrase+Arc</th>
</tr>
</thead>
<tbody>
<tr>
<td>SURE</td>
<td>65.60%</td>
<td>75.10%</td>
<td>76.30%</td>
</tr>
<tr>
<td>SURE + POSSIBLE</td>
<td>62.57%</td>
<td>78.79%</td>
<td>80.92%</td>
</tr>
</tbody>
</table>

All gains significant under Wilcoxon’s signed rank test

Results

- **SURE**
- **POSSIBLE**
- **Sure + Possible**

Learning

- Structured perceptron with averaging (Collins, 2002) to learn weights
- Phrase edit features \( \Phi(y) \) from MANLI (MacCartney et al., 2008): phrase sizes, lexical + contextual similarity, relative positions, etc
- Arc edit features \( \Phi(z) \) note the label category (e.g., subj) of the dependencies

Example: token alignments with syntactic support **missed** by the Meteor aligner (Denkowski & Lavie, 2011)

**Can we improve alignment using syntax?**

http://www.ling.ohio-state.edu/~mwhite/data/coling12/