A Joint Syntactic-Semantic Representation for Recognizing Textual Relatedness

Rui Wang
Saarland University, Germany

Yi Zhang & Günter Neumann
DFKI GmbH, Germany
RTE from 2-Way to 3-Way

- From RTE-3 pilot task
  - YES → Entailment
  - NO → Contradiction / Unknown

- Performance
  - RTE-4 (3-way): 0.51
  - RTE-4 (2-way): 0.57
  - RTE-3 (2-Way): 0.61
An Example

+ Text: At least five people have been killed in a head-on train collision in north-eastern France, while others are still trapped in the wreckage. All the victims are adults.

+ Hypothesis: A French train crash killed children.
An Example

+ Text: At least five people have been killed in a head-on train collision in north-eastern France, while others are still trapped in the wreckage. All the victims are adults.

+ Hypothesis: A French train crash killed children.

+ Contradictory but Related!
Entailment vs. Relatedness

- Textual Entailment
  - Unidirectional
  - Meaning preserved
  - Entailment vs. Non-entailment

- Textual Relatedness
  - Bidirectional
  - Weaker than similarity and stronger than co-occurrence
  - Related vs. Non-related (Unknown)
Strategies for 3-Way RTE

+ Traditional 2-way classification
  + Split E cases first: ECU → E/CU

+ Contradiction recognition (de Marneffe et al., 2008)
  + Split C cases first: ECU → C/EU

+ Others
  + Three-way classification: ECU → E/C/U
  + Split U cases first: ECU → U/EC
Strategies for 3-Way RTE

+ Traditional 2-way classification
  + Split E cases first: ECU → E/CU

+ Contradiction recognition (de Marneffe et al., 2008)
  + Split C cases first: ECU → C/EU

+ Others
  + Three-way classification: ECU → E/C/U
  + Split U cases first: ECU → U/EC
Baseline

- RTE-4 dataset
  - 500 E, 150 C, 350 U
  - NaiveBayes classifier, 10-fold CV
  - BoW + SynDep features (Wang and Neumann, 2007)

<table>
<thead>
<tr>
<th>Three-Way</th>
<th>Two-Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E/C/U$</td>
<td>$E/CU \rightarrow E/C/U$</td>
</tr>
<tr>
<td>53.20%</td>
<td>50.00%</td>
</tr>
<tr>
<td>/</td>
<td>82.80%</td>
</tr>
</tbody>
</table>
Outline

+ Recognizing Textual Relatedness
  + Related Work
  + Definition

+ The Joint Representation
  + Syntactic and Semantic dependency
  + Co-reference

+ Experiments & Results

+ Conclusion & Future Work
RTE vs. RTR

- RTE
  - Direct three-way classification (e.g. Agichtein et al., 2009); different rules simultaneously (Clark and Harrison, 2009)
  - Contradiction recognition (de Marneffe et al., 2008)

- Alignment
  - Phrased-based and dependency-graph-based (Pado et al., 2009)
  - Ontology-based (Siblini and Kosseim, 2009)
  - Dependency-path-based (Wang and Neumann, 2007)
Textual Relatedness

- Wang and Zhang (2009)
  - *If H is fully relevant to part of T, H is semantically related to T.*

- Relatedness
  - (Weaker than) Similarity
    - Surface string, semantic, etc.
  - (Stronger than) Co-occurrence
    - Distributionally or ontologically
Relationship between Relations

- CONTRADICTION
- ENTAILMENT
- Contradiction
- Consistency
- UNKNOWN

Relatedness

- Relatedness
- Relatedness

TAC 2009, Gaithersburg, MD, USA
Recognizing Textual Relatedness

+ Preprocessing

+ Dependency Parsing (MSTParser – McDonald et al. (2005))

+ Semantic Role Labeling (Zhang et al., 2008)
  + The CoNLL shared task (2008, 2009): 70~80%

+ Co-reference Resolution (BART – Versley et al. (2008))
Syntactic and Semantic Dependency
The Joint Representation
Decomposition of the Joint Graph
Decomposition (cont.)
Equations

\[ R(T, H) = \max_{1 \leq i \leq r, 1 \leq j \leq s} \{ R(\text{Tree}_{T_i}, \text{Tree}_{H_j}) \} \]

\[ R(\text{Tree}_T, \text{Tree}_H) = \min_{1 \leq i \leq n, 1 \leq j \leq m} \{ R(\langle P_T, D_{T_i}, A_{T_i} \rangle, \langle P_H, D_{H_j}, A_{H_j} \rangle) \} \]

\[ R(\langle P_T, D_T, A_T \rangle, \langle P_H, D_H, A_H \rangle) = \begin{cases} 
\text{Full} & R(P_T, P_H) = R(D_T, D_H) = R(A_T, A_H) = 1 \\
\text{NotFull} & R(P_T, P_H) = R(D_T, D_H) = 1 \\
\text{Other} & \text{Otherwise} \end{cases} \]
Lexical Semantic Resources

+ String matching of lemmas

+ Predicate
  + VerbOcean (Chklovski and Pantel, 2004)
  + Normalized Google Distance (NGD) (Cilibrasi and Vitanyi, 2007)

+ Argument
  + WordNet: synonym, hypernym, hyponym, antonym
  + NGD (available online)
Experiments

+ Run1
  + Wang and Zhang’s system + a backup using features from BoW and syntactic dependency

+ Run2
  + The main system (lenient) + a backup using features from BoW, syntactic, and semantic dependency

+ Run3
  + The main system (strict) + a backup using features from BoW and joint representation
## Results

<table>
<thead>
<tr>
<th>Runs</th>
<th>Main</th>
<th>Main -VO</th>
<th>Main -WN</th>
<th>Main -VO-WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>DFKI1</td>
<td>50.7%</td>
<td>50.5%</td>
<td>50.7%</td>
<td>50.5%</td>
</tr>
<tr>
<td>DFKI2</td>
<td>63.7%</td>
<td>63.2%</td>
<td>63.3%</td>
<td>63.0%</td>
</tr>
<tr>
<td>DFKI3</td>
<td>63.5%</td>
<td>63.3%</td>
<td>63.3%</td>
<td>63.3%</td>
</tr>
<tr>
<td>RTE-3</td>
<td>53.69%</td>
<td>53.19%</td>
<td>53.50%</td>
<td>52.88%</td>
</tr>
<tr>
<td>RTE-4</td>
<td>56.60%</td>
<td>56.00%</td>
<td>56.10%</td>
<td>55.7%</td>
</tr>
</tbody>
</table>
Results (cont.)

<table>
<thead>
<tr>
<th>System</th>
<th>DFKI2</th>
<th>Gold-Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>E</td>
</tr>
<tr>
<td>E</td>
<td></td>
<td>238</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>U</td>
<td></td>
<td>58</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>300</td>
</tr>
</tbody>
</table>
## Results (cont.)

<table>
<thead>
<tr>
<th>Runs</th>
<th>Main</th>
<th>Main -VO</th>
<th>Main -WN</th>
<th>Main -VO-WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>DFKI1</td>
<td>62.5%</td>
<td>62.5%</td>
<td>62.7%</td>
<td>62.5%</td>
</tr>
<tr>
<td>DFKI2</td>
<td>66.8%</td>
<td>66.5%</td>
<td>66.7%</td>
<td>66.3%</td>
</tr>
<tr>
<td>DFKI3</td>
<td>68.5%</td>
<td>68.3%</td>
<td>68.3%</td>
<td>68.3%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Runs</th>
<th>Main</th>
<th>Main -VO</th>
<th>Main -WN</th>
<th>Main -VO-WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>DFKI1</td>
<td>74.0%</td>
<td>73.7%</td>
<td>73.8%</td>
<td>73.7%</td>
</tr>
<tr>
<td>DFKI2</td>
<td>74.3%</td>
<td>73.7%</td>
<td>73.8%</td>
<td>73.5%</td>
</tr>
<tr>
<td>DFKI3</td>
<td>72.3%</td>
<td>72.2%</td>
<td>72.2%</td>
<td>72.2%</td>
</tr>
</tbody>
</table>
Summary

+ **Strategy**
  + 2-stage binary classification for 3-way RTE

+ **Approach**
  + Textual relatedness
  + Use a joint representation measure it

+ **Result**
  + Improved (combination)
  + Lexical resources
Future Work

- Two styles of alignment
  - Predicate (Dinu and Wang, 2009)
  - Argument (paraphrase resources?)

- Entailment vs. Contradiction
  - Fine-grained RTE
  - Specialized RTE modules

- Named-Entity vs. common nouns
Thank you!

Questions?

Or later