Intrinsic and Extrinsic Approaches to Recognizing Textual Entailment

Rui Wang
Intrinsic vs. Extrinsic
Intrinsic vs. Extrinsic

- Other Problem
- Other Problem
- Problem
- Other Problem
- Other Problem
- Other Problem
Recognizing Textual Entailment (RTE)

- Textual Entailment (Chierchia and McConnell-Ginet, 2000)
  - One text can be inferred by another text.

- An example
  - Text (T): Google files for its long awaited IPO.
  - Hypothesis (H): Google goes public.
Scope

• Logic Entailment
  – H is true OR T is not true

• Linguistic Implication
  – Conventional Implicature
  – Conversational Implicature

• Modality
Machine Translation (MT) Triangle

MT Triangle

Meaning

Language Understanding → Language Generation

Source Text → Direct Translation → Target Text

... → Semantic Parsing
    → Syntactic Parsing
    → POS Tagging
    → Tokenization
    → Sentence String
Outline

Corpora

TSR corpus (Wang and Sporleder, 2010)
AMT corpus (Wang and Callison-Burch, 2010)

(Linguistic) Preprocessing

Parsing (Zhang and Wang, 2009)
Semantic Role Labeling (Zhang et al., 2008; Zhang et al., 2009)

RTE System

TACTE system (Wang and Zhang, 2008)
TAC 2008 (Wang and Neumann, 2009)
DIRT (Dinu and Wang, 2009)
Relatedness recognition (Wang and Zhang, 2009)
TAC 2009 (Wang et al., 2009)
TSR recognition (Wang and Zhang, 2011)

Applications

Answer Validation (Wang and Neumann, 2008a)
Relation Validation (Wang and Neumann, 2008b)
PETE (Wang and Zhang, 2010)
Outline (cont.)

• **Intrinsic** Approaches
  – Architecture
  – RTE with Event Triples
  – RTE with Inference Rules

• **Extrinsic** Approaches
  – Motivation
  – Multi-Dimensional Classification Model

• **Summary and Perspectives**
Intrinsic Approaches
An Example

• **T**: *Bush used his weekly radio address to try to build support for his plan to allow workers to divert part of their Social Security payroll taxes into private investment accounts.*

• **H**: *Mr. Bush is proposing that workers be allowed to divert their payroll taxes into private accounts.*
Performance of the Existing Systems

![Graph showing performance metrics for RTE-3, RTE-4, and RTE-5 systems.](Image)
Architecture

RTE Corpora

Preprocessing

Parsing
Anaphora Resolution

NE Recognition
WSD

Entailment Recognition

YES / NO
Architecture

RTE Corpora

Preprocessing

RTE Module
- Candidate Identifier
- Entailment Recognition

RTE Module
- Candidate Identifier
- Entailment Recognition

RTE Module
- Candidate Identifier
- Entailment Recognition

RTE Module
- Candidate Identifier
- Entailment Recognition

Voting

YES / NO
Specialized RTE Module

• Two requirements
  – A good target
  – A good tackle

• Two examples
  – Event tuples containing named-entities
  – Tree skeleton with inference rules
Event Tuple

• < Event type, Time, Location, List<participants> >
  • Persons
  • Organizations

Rui Wang's PhD Defense
An Example

• **T:** *Released in 1995, Tyson returned to boxing, winning the World Boxing Council title in 1996. The same year, however, he lost to Evander Holyfield, and in a 1997 rematch bit Holyfield’s ear, for which he was temporarily banned from boxing.*

• **H:** *In 1996 Mike Tyson bit Holyfield’s ear.*
Event Time Pair

• T:
  – <released, 1995>
  – <winning, 1996>
  – <rematch, 1997>
  – <bit, 1997>

• H:
  – <bit, 1996>
Event Tuple

• **T:**
  – <released, 1995, N/A, <Tyson>>
  – <winning, 1996, N/A, <Tyson>>
  – <rematch, 1997, N/A, N/A>
  – <bit, 1997, N/A, N/A>

• **H:**
  – <bit, 1996, N/A, <Mike Tyson>>
Entailment Recognition

• Event type comparison
  – Lexical resources, e.g. WordNet, VerbOcean

• Named-entity comparison
  – Time expressions normalization and anchoring
  – Ontology of geographical terms
  – Person name / organization name partial matching
Inference Rules

• DIRT collection (Lin and Pantel, 2001)

\[ X \leftarrow \text{subj} - \text{prevent} - \text{obj} \rightarrow Y \]

\[ X \leftarrow \text{subj} - \text{provide} - \text{obj} \rightarrow \text{protection} - \text{mod} \rightarrow \text{against} - \text{pcomp} \rightarrow Y \]
T: Doctor Robin Warren and Barry Marshall received Nobel Prize ...

H: Robin Warren was awarded a Nobel Prize.
## Results

<table>
<thead>
<tr>
<th>Systems</th>
<th>RTE-2</th>
<th>RTE-3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BoW</td>
<td>TACTE</td>
</tr>
<tr>
<td>Coverage</td>
<td>100%</td>
<td>10.2%</td>
</tr>
<tr>
<td>Accuracy*</td>
<td>0.579</td>
<td><strong>0.684</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Systems</th>
<th>RTE-3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BoW</td>
<td>TACTE</td>
</tr>
<tr>
<td>Coverage</td>
<td>100%</td>
<td>8.1%</td>
</tr>
<tr>
<td>Accuracy*</td>
<td>0.611</td>
<td>0.655</td>
</tr>
</tbody>
</table>

*On covered data*
## Summary

<table>
<thead>
<tr>
<th>Target</th>
<th>Tackle</th>
<th>Preferences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event tuples</td>
<td>NE recognition</td>
<td>NE relation/resolution</td>
</tr>
<tr>
<td>Tree skeleton with inference rules</td>
<td>Matched by DIRT rules</td>
<td>Match tree skeleton with DIRT rules</td>
</tr>
</tbody>
</table>

### Future Extensions

<table>
<thead>
<tr>
<th>Negation and modality</th>
<th>Negation words or modal verbs</th>
<th>Scope and entailment rules</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logic inferencer</td>
<td>High confidence</td>
<td>Theorem prover</td>
<td>N/A</td>
</tr>
<tr>
<td>External knowledge bases</td>
<td>Covered cases</td>
<td>Knowledge application</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Extrinsic Approaches
Relevance and Necessity

things unrelated to the goal

No “necessity”

things related to the goal

things entailed by the goal

the goal
An Example

• **T:** 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**.

• **H:** A French train crash killed **children**.

• Contradictory but relevant!
Textual Semantic Relation (TSR)

- Contradiction
- Entailment
- Paraphrase
- Overlapping
- Unknown

Entailment

Unknown
Related Tasks

• Contradiction recognition
  – *Contradiction vs. Others* (de Marneffe et al., 2008)

• Paraphrase acquisition
  – *Paraphrase vs. Others* (a survey by Androutsopoulos and Malakasiotis (2010))

• Directionality recognition
  – *Entailment vs. Paraphrase* (e.g., Kotlerman et al., 2009)
RTE

RTE Rectangle

Meaning \( (T) \) \( \rightarrow \) Meaning \( (H) \)

Language Understanding \( \rightarrow \) Language Understanding

Text \( (T) \) \( \rightarrow \) Hypothesis \( (H) \)

Simplification \( \rightarrow \) Simplification

Representation \( (T) \) \( \rightarrow \) Representation \( (H) \)

TSR Rectangle

Meaning \( (T_1) \) \( \rightarrow \) Meaning \( (T_2) \)

Language Understanding \( \rightarrow \) Language Understanding

Text \( (T_1) \) \( \rightarrow \) Text \( (T_2) \)

Simplification \( \rightarrow \) Simplification

Representation \( (T_1) \) \( \rightarrow \) Representation \( (T_2) \)
Meaning Representation

• The **TACTE** system: Event time pairs
• The **ExTACTE** system: Event tuples
• The **ExDIRT** system: Tree skeletons

• Dependency triple sets: \{DEP_T\} and \{DEP_H\}
  – Syntactic dependency tree
  – Semantic dependency graph
  – Joint Dependency graph
Meaning Representation (cont.)

• **H**: *Value is questioned.*

• Syntactic dependency
  – `<is, SBJ, value>`
  – `<is, VC, questioned>`

• Semantic dependency
  – `<questioned, A1, value>`
Meaning Representation (cont.)

- \textbf{T}: Devotees of the market question the value of the work.
Direct Classification

<\text{T}, \text{H}>

? 

Paraphrase (P) 
Entailment (E) 
Contradiction (C) 
Unknown (U)
Two-Stage Classification

$<T, H> \rightarrow \text{Relatedness}$

$\text{Paraphrase (P)} \rightarrow \text{?} \rightarrow \text{Entailment (E)} \rightarrow \text{Contradiction (C)} \rightarrow \text{Unknown (U)}$

3.4% ↑
Performance of the Existing Systems

- Three-way RTE: *Entailment*, *Contradiction*, *Unknown*
The 3-Dimensional Model

\(<T, H>\)

- Relatedness
- Consistency
- Equivalence

- Paraphrase (P)
- Entailment (E)
- Contradiction (C)
- Unknown (U)
The 3-Dimensional Model (cont.)

- Two-stage classification
- Three measurements
  - Relatedness
  - Consistency
  - Equivalence

<table>
<thead>
<tr>
<th></th>
<th>Relatedness</th>
<th>Consistency</th>
<th>Equivalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paraphrase (P)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Entailment (E)</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Contradiction (C)</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Unknown (U)</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>
## Datasets

<table>
<thead>
<tr>
<th>Corpora</th>
<th>Paraphrase (P)</th>
<th>Entailment (E)</th>
<th>Contradiction (C)</th>
<th>Unknown (U)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMT (584)</td>
<td>/</td>
<td>Facts (406)</td>
<td>Counter-facts (178)</td>
<td>/</td>
</tr>
<tr>
<td>MSR (5841)</td>
<td>Paraphrase (3940)</td>
<td>Non-Paraphrase (1901)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PETE (367)</td>
<td>/</td>
<td>YES (194)</td>
<td>NO (173)</td>
<td></td>
</tr>
<tr>
<td>RTE (2200)</td>
<td>ENTAILMENT (1100)</td>
<td>CONTRADICTION (330)</td>
<td>UNKNOWN (770)</td>
<td></td>
</tr>
<tr>
<td>Total (9252)</td>
<td>3943</td>
<td>637</td>
<td>525</td>
<td>973</td>
</tr>
</tbody>
</table>
## Results (RTE)

<table>
<thead>
<tr>
<th>Systems</th>
<th>4-Way</th>
<th>3-Way</th>
<th>2-Way</th>
</tr>
</thead>
<tbody>
<tr>
<td>(C, E, P, U)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct BoW</td>
<td>39.3%</td>
<td>54.5%</td>
<td>63.2%</td>
</tr>
<tr>
<td>Direct Joint</td>
<td>42.3%</td>
<td>50.9%</td>
<td>66.8%</td>
</tr>
<tr>
<td>Only Relatedness (Our Prev.)</td>
<td>/</td>
<td>59.1%</td>
<td>69.2%</td>
</tr>
<tr>
<td>3-D Model</td>
<td>45.9%</td>
<td>58.2%</td>
<td>69.9%</td>
</tr>
<tr>
<td>MacCartney and Manning (2007)*</td>
<td>/</td>
<td>/</td>
<td>59.4%</td>
</tr>
<tr>
<td>Heilman and Smith (2010)*</td>
<td>/</td>
<td>/</td>
<td>62.8%</td>
</tr>
</tbody>
</table>

*Different test data*
## Results (Paraphrase)

<table>
<thead>
<tr>
<th>P vs. C&amp;E&amp;U</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-D Model</td>
<td>79.6%</td>
<td>57.2%</td>
<td>72.8%</td>
</tr>
<tr>
<td>Das and Smith (2009) (QG)*</td>
<td>73.9%</td>
<td>74.9%</td>
<td>91.3%</td>
</tr>
<tr>
<td>Das and Smith (2009) (PoE)*</td>
<td>76.1%</td>
<td>79.6%</td>
<td>86.0%</td>
</tr>
<tr>
<td>Heilman and Smith (2010)*</td>
<td>73.2%</td>
<td>75.7%</td>
<td>87.8%</td>
</tr>
</tbody>
</table>

*Different test data*
Results (cont.)
Summary and Perspectives
Intrinsic Approaches

• Event tuple
  – Event time pair
  – Extended to other slots (NEs)

• (Textual) inference rules
  – DIRT (with tree skeleton)
  – More general paraphrase resources (ongoing)
Extrinsic Approaches

• Textual semantic relations
  – Relatedness recognition
  – Extended with two other dimensions

• Specialized TSR modules
  – Split the data (as for RTE)
  – Systematic feature engineering
Thank YOU!