Efficient Semantic Deduction and Approximate Matching over Compact Parse Forests

Roy Bar-Haim\textsuperscript{1}, Jonathan Berant\textsuperscript{2}, Ido Dagan\textsuperscript{1}, Iddo Greental\textsuperscript{2}, Shachar Mirkin\textsuperscript{1}, Eyal Shnarch\textsuperscript{1}, Idan Szpektor\textsuperscript{1}

\textsuperscript{1}Bar-Ilan University
\textsuperscript{2}Tel-Aviv University
Our RTE-3 Recipe

1. Represent $t$ and $h$ as parse trees
2. Try to prove $h$ from $t$ based on available knowledge
3. Measure the “distance” from $h$ to the generated consequents of $t$.
4. Determine entailment based on distance
5. Cross your fingers …

*I forgot one thing…*

6. **Wait a few hours…**

*This year: hours $\rightarrow$ minutes*
Textual Entailment – Where do we want to get to?

- Long term goal: **robust semantic inference engine**
  - To be used as a generic component in text understanding applications
  - Encapsulating all required inferences
- Based on compact, well defined formalism:
  - Knowledge representation
  - Inference mechanisms
Our Inference Formalism
Bar-Haim et al., AAAI-07 & RTE3

- A *proof system* over parse trees:
  - Represents diverse kinds of semantic knowledge uniformly as entailment (inference) rules
  - Allows unified inference mechanism

- Analogous to logic proof systems:

<table>
<thead>
<tr>
<th>Propositions</th>
<th>Parse Trees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inference Rules</td>
<td>Tree transformations</td>
</tr>
<tr>
<td>Proof</td>
<td>A sequence of trees, generated by rule application</td>
</tr>
</tbody>
</table>

- Given *Text (T)* and *Hypothesis (H)*, try to *generate H* from *T*
- Formalizes common transformation-based approaches
A Sample Proof

Consequent

⇒ Barack Obama is the elected
⇒ Barack Obama won the elections
⇒ The Democratic nominee won the elections.

Hypothesis: The Democratic nominee won the elections.

Text: McCain congratulated the elected president, Barack Obama.

Entailment Rule

Apposition (syntactic)

X is the elected president
⇒ X won the elections
⇒ Barack Obama won the elections
⇒ The Democratic nominee won the elections.
Explicit Generation of Consequents

Text: Children like candies.

Rules: children $\rightarrow$ kids ; like $\rightarrow$ enjoy ; candies $\rightarrow$ sweets

Consequents:
Kids like candies.
Kids enjoy candies.
Children like sweets.
...
$2^3$ alternatives!

We need compact representation of consequents!
Intuitive Solution: Add only the Entailed Part

- Resulting structure is a union of sentences (trees)
- Each of its sub-structures is assumed to be entailed
Intuitive Direction: Add only the Entailed Part

- Resulting structure is a union of sentences (trees)
- Each of its sub-structures is assumed to be entailed

![Diagram]

Semantics not preserved!
How to share common variables?

The resigned $X \rightarrow$ the $X$ has resigned

How to combine these structures efficiently (and correctly)?
Talk Outline

- A novel data structure, *compact forest*, for efficient inference
  - Preserves semantics
- New sources of entailment rules
  - Scale made feasible by compact forest
- Approximate matching
- Results
The Compact Forest
A sample rule: passive-to active transformation

The book was read by John yesterday
John read the book yesterday
The book was read by John yesterday.
The book was read by John yesterday.
The book was read by John yesterday → N1 \ V2

Step 3
Connect R as alternative to L

Disjunction edges (d-edges)
Specify disjoint alternatives for source and target nodes

V

N1
N2

obj

subj

by

pcomp-n

i

read

by

book

det

the

mod

ROOT

be

The book was read by John yesterday

The book was read by John yesterday

N1 \ V2
The book was read by John yesterday ⇒ N1 V N2
The book was read by John yesterday ⇒ N1 read N2

Step 4
instantiation of V

V has different modifiers in L and R ⇒ should be copied

The book was read by John yesterday ⇒ N1 read N2
The book was read by John yesterday ⇒ N1 read N2
The book was read by John yesterday ⇒ John read the book
The book was read by John yesterday ⇒ John read the book
The book was read by John yesterday ⇒ John read the book yesterday
The book was read by John yesterday.

An efficient data structure, preserving inference semantics.
Children and Sweets – the Compact Version…

Complexity reduction from exponential to linear!
New Knowledge Sources
Extracting Lexical Rules from Wikipedia

E.T. the Extra-Terrestrial

From Wikipedia, the free encyclopedia
(Redirected from E.T. (film))

*E.T. the Extra-Terrestrial* is a 1982 science fiction film co-produced and directed by Steven Spielberg, written by Melissa Mathison and starring Henry Thomas, Robert MacNaughton, Drew Barrymore, Dee Wallace and Peter Coyote. It tells the story of Elliott (played by Thomas), a lonely boy who befriends a friendly alien dubbed "E.T." who is stranded on Earth. Elliott and his family, including his sister Gertie (played by Barrymore), her pet dog "Buck," and their mother Mary (played by Wallace), work to help E.T. return to his home planet.

- **Be-complement:** Nominal complements of ‘be’
- **Redirect:** various terms to canonical title
- **Parenthesis:** used for disambiguation
- **Link:** Maps to a title of another article
Lexical-Syntactic Rules from Lexical Resources

- Entailment rules between predicates + argument mapping
- Combining information from various lexical resources:
  - **WordNet**: semantic relations (synonyms, hyponyms)
  - **VerbNet**: subcategorization frames for verbs
  - **Nomlex-Plus**:
    - Mapping between verbs and nominalizations (acquire ⇔ acquisition)
    - Subcategorization frames for nominalizations
  - Example:
    
    \[ \text{homicide of } X \text{ by } Y \Leftrightarrow \text{killing of } X \text{ by } Y \Leftrightarrow Y \text{ kill } X \]
Polarity Rules

- Annotate **polarity** for verbs and Nouns
  - **Positive**: John called\(^{(+)}\) Mary.
  - **Negative**: John forgot\(^{(-)}\) to call Mary.
  - **Unknown**: John wanted to call\(^{(?)}\) Mary.

- Expressed by
  - Verbal negation
  - modal verbs and adverbials
  - conditional sentences
  - Verbs inducing **negative** polarity - forget, failed
  - Verbs inducing **unknown** polarity - want, attempt

- Polarity mismatch is indicative for no-entailment
  - Used as feature
Adding Approximate Matching
RTE4 System Architecture

**Rule Bases**
- Wikipedia
- WordNet (+arguments)
- Polarity
- Syntactic
- DIRT
- WordNet (lexical)

**New rule bases**
- RTE3 rule bases

**Preprocessing**
- parsing, co-reference, NER, number normalization

**Knowledge-based Inference**
- rule application

**Approximate Matching**
- Feature Extraction and Classification
Features for Approximate Matching

- Coverage of H by F
  - Lexical coverage (words, verbs, numbers, named entities)
  - Local syntactic coverage (edges)
  - Global structural matching
    - Aim to match maximal subtrees of H in F
    - Efficient computation using dynamic programming
- Polarity mismatch (*forgot to buy* vs. *bought*)
- Argument matching for corresponding predicates
Results

- Using SVM classifier (outperforms decision tree)

<table>
<thead>
<tr>
<th>Train</th>
<th>Test</th>
<th>Knowledge-Based Inference</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTE3 Dev+Test</td>
<td>RTE4</td>
<td>✓</td>
<td>58.4</td>
</tr>
<tr>
<td>RTE3 Dev</td>
<td>RTE4</td>
<td>✓</td>
<td>60.3</td>
</tr>
<tr>
<td>RTE3 Dev</td>
<td>RTE3 Test</td>
<td>✓</td>
<td>66.9</td>
</tr>
<tr>
<td>RTE3 Dev</td>
<td>RTE3 Test</td>
<td></td>
<td>64.6</td>
</tr>
<tr>
<td>RTE3 Dev</td>
<td>½RTE4</td>
<td>✓</td>
<td>58.8</td>
</tr>
<tr>
<td>RTE3 Dev</td>
<td>½RTE4</td>
<td></td>
<td>55.8</td>
</tr>
</tbody>
</table>

- Competitive accuracy on RTE-3 (~4-6 out of 26)
- Knowledge-based inference improves accuracy (+3%)
- New efficient architecture reduced running time from hours to minutes, allowing application of many more rules
Conclusion

- **Takeouts**
  - A formalized proof system over parse trees can be implemented *efficiently*
    - While preserving semantics – equivalent to explicit generation of consequents
    - Substantial run time reduction
    - Substantial increase in number of rules applied
  - Knowledge helps!
    - But not much for now...

- **Future research**
  - Analysis (errors, missing knowledge, impact of resources, search strategies)
  - Additional knowledge sources
  - Improve approximate matching

Thank You!