Multiple Alternative Sentence Compressions and Word-Pair Antonymy for Automatic Text Summarization and Recognizing Textual Entailment

Saif Mohammad[1], Bonnie Dorr[1,2], Melissa Egan[1], Jimmy Lin[1], Nitin Madnani[1], David Zajic[3]
University of Maryland[1], HLT Center of Excellence[2], CASL[3]

Special thanks to Marie-Catherine de Marneffe, Sebastian Pado, Christopher Manning, and the Stanford University RTE team.

Antonymy

Kinds

- Clear opposites:
  - *wet – dry*
  - *hard – soft*
  - *promoted – demoted*

→ 
manually created lexicons

- Contrasting word pairs:
  - *promoted – censured*
  - *hard – fluid*
  - *flinch – advance*

→ 
largely unrecorded

Computing word-pair antonymy: Mohammad et al. 2008

Method:
- Identify contrasting word pairs
  - using seed antonym pairs and thesaurus categories.
- Determine degree of antonymy
  - using distributional distance and tendency to co-occur.

Evaluation:
950 GRE-style closest-opposite questions.

Results:
F score = .70 (baselines: .20 and .22).

Why be antonymy-aware

- **Recognizing Textual Entailment and Contradictions**:
  Antonyms and polarity can preserve or contradict meaning.
  - **paraphrases**:
    Sirius Black could not evade the dementors.
    The dementors caught Sirius Black.
  - **disagreement, contention, and contradiction**:
    Giuliani's 9/11 emergency management was prompt.
    Slow response was one of his biggest criticisms.
- **Summarization:**
  Presence of antonyms is an indicator of summary-worthy information.
  - **contrast**:
    Gregory Peck can play both strong and sensitive roles.
    Peck can only play simple roles, not complex ones.
  - identifying entailments and contradictions
- Sentiment detection, detecting humor, improving distributional thesauri.

Recognizing Textual Entailment and Contradictions

Entailment and contradiction

A practical definition of **entailment**:
A text (or source) T entails a hypothesis H if a normal reader would be happy to accept T as strong evidence that H is true (assuming that T is reliable).

A practical definition of **contradiction**:
T contradicts H if it is very unlikely that both T and H can be true at the same time.

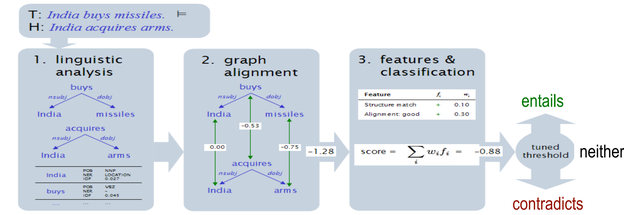
Example:

T1 Internet media company Yahoo Inc. announced Monday it is buying Overture Services Inc. in a $1.63-billion (U.S.) cash-and-stock deal that will bolster its on-line search capabilities.

Entails:
H1.1 *Yahoo bought Overture*
H1.2 *Overture was acquired by Yahoo*
H1.3 *Overture was bought*
H1.4 *Yahoo is an internet company*

Contradicts:
H1.5 *Overture bought Yahoo*
H1.6 *Yahoo sold Overture*

Stanford three-stage RTE system



Antonym features

- Check if an aligned pair of words (across T and H) are antonyms.
- If yes, then boolean features generated:
  - Antonyms appear in contexts of matching polarity
    Hitler survived the plane crash.
    Hitler died in a plane crash
  - Only the hypothesis (or source) word appears in negative polarity context
    Hitler survived the plane crash.
    Hitler did not die in a plane crash.

Summarization

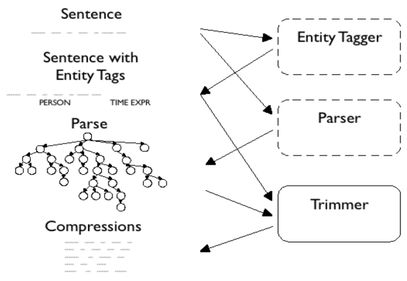
UMD three-stage architecture for summarization:

1. Tagging: the Stanford Parser (Klein and Manning 2003).
- Sentences are part-of-speech tagged and parsed.
- Named entities in the sentences are identified.

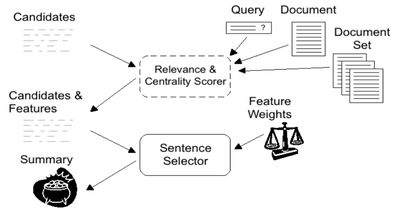
2. Sentence compression: Trimmer (Zajic 2007)
- Linguistically motivated rules to mask syntactic components of the parse of a source sentence.
- Rules are applied iteratively, and in many combinations
- Compression-specific feature values are assigned:
  - number of rule applications
  - parse tree depth of various rule applications.

3. Candidate selection
- Static features:
  - position of sentence in the document; length; compression-specific features; relevance scores
- Dynamic features include:
  - redundancy with current summary state; number of candidates from the same source document already in the summary
- Candidates are selected for inclusion until the summary reaches prescribed word limit or the pool is exhausted.

Stages 1 and 2: tagging and sentence compression.



Stage 3: candidate selection.



Trimmer + Antonymy Features

- Examine each of the words in a sentence to determine whether it has an antonym within the same within the same document.
  - If not, then the antonymy score contributed by this word is 0.
  - If yes, then the antonymy score contributed by this word is its degree of antonymy with the word it is most antonymous to.
- The antonymy score of a sentence is the sum of scores of constituent words.

Document fragment:

…
For a long time, physicists thought that the Universe will eventually contract.
New calculations show that in fact the Universe will continue to expand.
…

TAC 2008: RTE task

Data

1000 source and hypothesis pairs

Three submissions by Stanford--UMD

WordNet : used only WordNet antonyms
Automatic: used antonyms determined from our automatic method
All : used both manually and automatically determined antonyms

Results

	WordNet	Automatic	All
2-way accuracy	61.7	61.7	61.7
3-way accuracy	55.4	55.6	55.6
Avg. precision	44.08	44.26	44.27

Conclusions

- Our system stood 7th among the 33 participating systems.
- Performance using automatically generated antonyms and those compiled from a manually created resource such as WordNet is comparable.
  - Appealing method for resource-poor languages.
- Using the manually generated antonyms in addition to the automatically generated antonyms did not improve performance by much.

Future work

- Apply the method to data richer in contradictions.
- Analyze the manifestation of antonyms in contradictory source—hypothesis pairs.
- Use sophisticated antonymy features that take into account syntactic dependencies.

TAC 2008: Opinion Summarization Task

Blog data

- 609 documents covering 25 topics
- Writing style is informal
- Natural language text is enmeshed in metadata.

Extracting text to be summarized

- Extract content from HTML <BODY> … </BODY> tags
- Decode HTML-encoded characters
  - e.g., " " for space, "&amp" for ampersand
- Convert HTML separator tags into newlines
  - e.g.,
, <HR>, <TD>, <P>
- Remove remaining HTML tags
- Remove common non-content phrases
  - e.g., "Posted by...", "Published by...", "Related Stories", "Track-Back", "Blog This", "Copyright"
- Filter out any line of text containing fewer than *n* = 6 words.

UMD submissions

1. Trimmer
2. Trimmer + Antonymy features

Results

	Trimmer		Trimmer + Antonymy	
	Rank	Score	Rank	Score
Pyramid	11/19	.14/1	13/19	0.130/1
Grammaticality	16/19	4.4/10	17/19	4.318/10
Non-redundancy	3/19	6.7/10	6/19	6.364/10
Coherence	13/19	2.4/5	11/19	2.409/5
Fluency	9/19	3.6/10	15/19	3.318/10

Conclusions

- The performance of the UMD summarizer was roughly middle-of-the-pack.
- It was particularly strong in non-redundancy (rank 3).
- Adding antonymy features:
  - improved coherence
  - negatively affected other aspects.

Future Work

- Use antonymy features more extensively
  - focus on antonyms in adjacent sentences;
  - include syntactic dependency information.

References

- Saif Mohammad, Bonnie Dorr, and Graeme Hirst. October 2008. Computing Word-Pair Antonymy. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2008)*, Waikiki, Hawaii.
- Sebastian Pado, Marie-Catherine de Marneffe, Bill MacCartney, Anna N. Rafferty, Eric Yeh, and Christopher D. Manning. November 2008. Manning. Deciding Entailment and Contradiction with Stochastic and Edit Distance-based Alignment. In *Proceedings of the Text Analysis Conference (TAC 2008)*, Gaithersburg, MD.
- David M. Zajic. 2007. Multiple Alternative Sentence Compressions (MASC) as a Tool for Automatic Summarization Tasks. Ph.D. thesis, University of Maryland, MD.

Acknowledgment

We thank Marie-Catherine de Marneffe, Sebastian Pado, Christopher Manning, and the rest of the Stanford University RTE team for working with us in putting forward a joint Stanford--UMD submission. This work was supported, in part, by the National Science Foundation under Grant No. IIS-0705832, in part, by the Human Language Technology Center of Excellence, and in part, by the Natural Sciences and Engineering Research Council of Canada. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsor.