Computing word-pair antonymy: Mohammad et al. 2008

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

Evaluation:
950 GALE-style closest-opposite questions.

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

Why be antonymy-aware:
- Recognizing Textual Entailment and Contradictions: Antonymy and polarity can preserve or contradict meaning.
- Paraphrases: Sirius Black could not evade the dementors. The dementors caught Sirius Black.
- Disagreement, contention, and contradiction: Guldan’s 9/11 emergency management was prompt. Slow response was one of his biggest criticisms.

Summary:
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 thesaurus.

Kinds
- Clear opposites: wet – dry
- Promoted – demoted
- Contrasting word pairs: promoted – rendered
- Hard – fluid
- Flinch – advance

Largely unrecorded

manually created lexicons

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:
T: 1/1 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

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
  - Using the manually generated antonym in addition to the automatically generated antonym 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 antonym features that take into account syntactic dependencies.

UMD three-stage architecture for summarization:

- 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:
  - 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.

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 antonym score contributed by this word is 0.
  - If yes, then the antonym score contributed by this word is the degree of antonymy of the word it is most antonymous to.
- The antonym score of a sentence is the sum of scores of constituent words.

MDM three-stage architecture for summarization:

- 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:
  - 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.

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 antonym score contributed by this word is 0.
  - If yes, then the antonym score contributed by this word is the degree of antonymy of the word it is most antonymous to.
- The antonym score of a sentence is the sum of scores of constituent words.

References
- Pado, Christopher Manning, and the rest of the Stanford University RTE team. In Proceedings of the Text Analysis Conference (TAC 2008), Gaithersburg, MD.
- Zajic, Gregory Peck can play both strong and sensitive roles. Peck can only play simple roles, not complex ones.

Future Work
- Use antonymy features more extensively:
  - Focus on antonyms in adjacent sentences
  - Include syntactic dependency information.

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.

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>"...
- Decode HTML-encoded characters:
  - e.g., "&nbsp;" for space, "&amp;" for ampersand
- Convert HTML separator tags into newlines:
  - e.g., "<BR>, "<HR>, "<TD>", etc.
- 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

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.