Finding Hedges by Chasing Weasels: Hedge Detection Using Wikipedia Tags and Shallow Linguistic Features

Viola Ganter and Michael Strube
EML Research gGmbH Heidelberg, Germany
ACL and AFNLP, 2009

22 May, 2014
Jaehwan Lee

Outline

- Introduction
- Related Work
- Weasel Words
- Data and Annotation
- Method
  - Words Preceding Weasel Tags
  - Adding shallow linguistic features
- Results and Discussion
- Conclusions

Introduction

- Distinguishing facts from fiction
- Indicate that speakers do not back up their opinions with facts
- in Abstract
  - “We investigate the automatic detection of sentences containing linguistic hedges using corpus statistics and syntactic patterns

Introduction

- Distinguishing facts from fiction
- Indicate that speakers do not back up their opinions with facts
- in Abstract
  - "We investigate the automatic detection of sentences containing linguistic hedges using **corpus statistics** and **syntactic patterns**

Related Work

- Focused on the biomedical domain
  - Light et al. (2004)

- Weakly supervised system for hedge classification
  - in a very narrow subdomain in the life sciences
  - Medlock and Briscoe (2007)

Weasel Words

- Wikipedia editors are advised to avoid *weasel words*
  - E.g. "Some people say …", "I think …", "Clearly …"

- Wikipedia style guidelines instruct editors to
  - if they notice weasel words, insert a {{weasel-inline}} or a {{weasel-word}} tag to mark sentences or phrases for improvement
  - E.g. Others argue {{weasel-inline}} that the news media are simply catering to public demand.

- Many Wikipedia articles contain a specific weasel tag
  - so that Wikipedia can be viewed as a **readily annotated corpus**

Data and Annotation

- Balanced set
  - chose one random, non-tagged sentence per tagged sentence
  - Wikipedia dumps from years 2006 to 2008
    - articles that contained the string {{weasel
    - 168,923 unique sentences containing 437 weasel tags
    - one random, non-tagged sentence per tagged sentence
    - resulting in a set of 500 sentences
  - Wikipedia dumps completed on March 6, 2009
    - 70,436 sentences with 328 weasel tags
    - Again, a balanced set of 500 sentences
- Manually annotated set
  - expected there to be a much higher number of potential weasel words which had not yet been tagged leading to false positives
  - one of the authors, two linguists and one computer scientist
  - resulting in a set of 246 sentences for evaluation

Method

- in Abstract
  - “We investigate the automatic detection of sentences containing linguistic hedges using **corpus statistics and syntactic patterns**
- Corpus statistics
  - Words Preceding Weasel Tags (wpw)
- Syntactic patterns
  - Adding shallow linguistic features (asp)

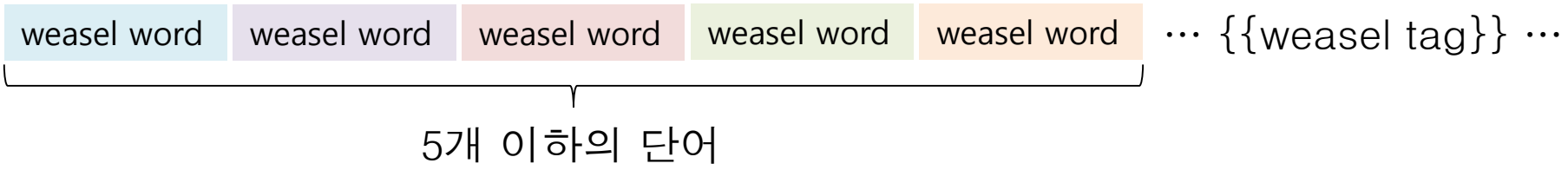
Words Preceding Weasel Tags (wpw)

- Assumption
  - weasel phrases contain at most five words
  - weasel tags are mostly inserted behind weasel words or phrase

… {{weasel tag}} …

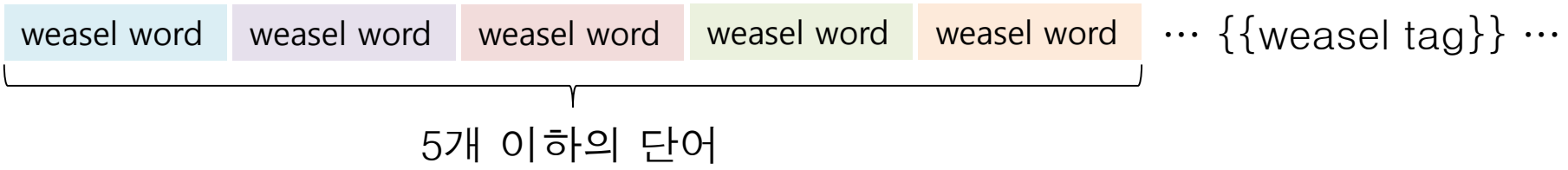
Words Preceding Weasel Tags (wpw)

- Assumption
  - weasel phrases contain at most five words
  - weasel tags are mostly inserted behind weasel words or phrase



Words Preceding Weasel Tags (wpw)

- Assumption
  - weasel phrases contain at most five words
  - weasel tags are mostly inserted behind weasel words or phrase



- Two Factors
  - Relative frequency
  - Average distance

- Equation goes,

$$Score(w) = RelF(w) + AvgDist(w) \qquad (1)$$

Words Preceding Weasel Tags (wpw)

- Relative frequency

$$RelF(w) = \frac{W(w)}{log_2(C(w))} \qquad (2)$$

  - W(w) : the number of times word w occurred in the context of a weasel tag
  - C(w) : the total number of times w occurred in the corpus
  - to give those words a high score which occur frequently in the context of a weasel tag

  - due to the sparseness of tagged instances, words that occur with a very high frequency in the corpus automatically receive a lower score than low-frequent words
  - Thus, use the logarithmic function to diminish this effect

Words Preceding Weasel Tags (wpw)

- Average distance

$$AvgDist(w) = \frac{W(w)}{\sum_{j=0}^{W(w)} dist(w, weaseltag_j)} \quad (3)$$

  - j : each weasel context
  - dist(w, $weaseltag_j$) : the distance of word w to the weasel tag in j

  - E.g. A word that always appears directly before the weasel tag will receive an AvgDist value of 1
  - E.g. A word that always appears five words before the weasel tag will receive an AvgDist value of 1/5

Words Preceding Weasel Tags (wpw)

- Normalization
  - wpw(S) : the sum of scores over all words in S
  - normalized by the hyperbolic tangent

$$wpw(S) = \tanh \sum_{i=0}^{|S|} Score(w_i) \qquad (5)$$

with $|S|$ = the number of words in the sentence.

- Classification
  - After calculating wpw(S) score for a sentence S
  - if wpw(S) is larger than a threshold, it is classified as weasel

$$S \rightarrow weasel \;\; if \;\; wpw(S) > \sigma \qquad (4)$$

Adding Shallow Linguistic Features (asp)

- the Weasel words in Wikipedia can be divided into
  - Numerically underspecified subjects ("Some people", "Many")
  - Passive constructions ("It is believed", "It is considered")
  - Adverbs ("Often", "Probably")

- If a pattern is found,
  - only the head of the pattern is assigned a score
  - i.e. adverbs, main verbs for passive patterns, nouns and quantifiers for numerically underspecified subjects

$$asp(S) = \tanh \sum_{i=0}^{heads_S} Score(w_i) \qquad (6)$$

where $heads_S$ = the number of pattern heads found in sentence S.

Results and Discussion

- Result
  - Both model perform comparably well on the development test data
  - the syntactic patterns do not contribute to the regeneration of weasel tags
  - Word frequency and distance to the weasel tag are sufficient

- Limitation
  - decreasing precision of both approaches when trained on more tagged sentences (i.e., computed with a higher threshold) might be caused by the great number of unannotated weasel words
  - A disadvantage of the weasel tag is its short life span

Results and Discussion

- Comparison
  - difference becomes more distinct when manually annotated data form the test set
  - *asp* out performs *wpw* by a large margin (also because *wpw* performs rather poorly)
  - suggests that the added syntactic patterns indeed manage to detect weasels that have not yet been tagged

σ	.60	.70	**.76**	.80	.90	**.98**
balanced set						
wpw	.68	.68	.68	.69	.69	**.70**
asp	.67	.68	**.68**	.68	.61	.59
manual annot.						
wpw	-	.59	-	-	-	**.59**
asp	.68	.69	**.69**	.69	.70	.65

Table 2: F-scores at different thresholds (bold at the precision/recall break-even-points determined on the development data)

Conclusions

- Main Idea
  - to use Wikipedia as a readily annotated corpus
- The experiments show that
  - the syntactic patterns work better on manual annotations
  - word frequency and distance work better on Wikipedia weasel tags itself
- This approach
  - takes a much broader domain than previous work
  - easily be applied to different languages by using Wikipedia
  - using the Wikipedia edit history will resolve short span of weasel tags