Word Epoch Disambiguation

Authors: Mihalcea and Nastase
Presented by Saheel

2012-09-14 Fri

Introduction

- Language – a constant?

Introduction

- Language – a constant?
  - Sense change, context change, new words, etc.
  - “run”, “gay”, “photography”

Introduction

- Language – a constant?
  - Sense change, context change, new words, etc.
  - “run”, “gay”, “photography”
- Study of Language as an evolutionary phenomenon

Introduction

- Language – a constant?
  - Sense change, context change, new words, etc.
  - “run”, “gay”, “photography”
- Study of Language as an evolutionary phenomenon
  - Requires large amounts of non-contemporary data
  - Google Books and Google Ngrams?

Introduction – contd.

- Quantify changes in word usage due to

Introduction – contd.

- Quantify changes in word usage due to
  - Sense change
  - Frequency change
  - Change in co-occuring topics
  - ...

Introduction – contd.

- Quantify changes in word usage due to
  - Sense change
  - Frequency change
  - Change in co-occuring topics
  - ...
- Using Google Books
  - 3 epochs: 1800+/-25 yrs, 1900+/-25 yrs, 2000+/-25 yrs

Related Work

- Previous Work at higher level

Related Work

- Previous Work at higher level
  - Culturomics – changes in human nature
  - Language = Genetics?
  - Scientific topics over time
  - Changes in context reflecting events occuring at that time
- How words themselves change?

WED formulation

WED formulation

- Classify period of time when a word was used

WED formulation

- Classify period of time when a word was used
- Epoch = Period of 50 years

WED formulation

- Classify period of time when a word was used
- Epoch = Period of 50 years
- Publication year of book $\sim$ Time when word is used

Experimental Setting – Target Words

- Driven by aim: Sense change or Context change?

Experimental Setting – Target Words

- Driven by aim: Sense change or Context change?
  - Polysemous and Monosemous words
  - Words frequent in all epochs and in one epoch

Experimental Setting – Target Words

- Driven by aim: Sense change or Context change?
  - Polysemous and Monosemous words
  - Words frequent in all epochs and in one epoch
- 50 words for each open class (nouns, adj, verbs, and adv)
  - 25 polysemous and 25 monosemous
  - 10-5-5-5 frequency-wise distribution
- 165 with enough examples

Experimental Setting – Data

- Top 100 snippets per target word per epoch
- Processing snippets
  - Tokenization
  - POS tagging
- POS Unambiguity

Experimental Setting – Example

1800: On reaching Mr. Crane's house, **dinner** was set before us ; but as is usual here in many places on the Sabbath, it was both **dinner** and tea combined into a single meal.

1900: The average **dinner** of today consists of relishes; of soup, either a consomme (clear soup) or a thick soup.

2000: Preparing **dinner** in a slow cooker is easy and convenient because the meal you're making requires little to no attention while it cooks.

Algorithm and Evalution

- Local and Topical features (similar to Ng and Lee, 1996)

Algorithm and Evalution

- Local and Topical features (similar to Ng and Lee, 1996)
- Naive Bayes Classifier (Ng and Lee, 2002)

Algorithm and Evalution

- Local and Topical features (similar to Ng and Lee, 1996)
- Naive Bayes Classifier (Ng and Lee, 2002)
- 10-fold Cross validation
- Baseline: Most frequent class

Results and Discussion

- Absolute Improvement of 18.5%

Results and Discussion

- Absolute Improvement of 18.5%
  - “computer” (42%) and “install” (41%)
  - “again” (3%) and “captivate” (7%)

Results and Discussion

- Absolute Improvement of 18.5%
  - "computer" (42%) and "install" (41%)
  - "again" (3%) and "captivate" (7%)
- 2 kinds of analysis:
  - Change in frequency
  - Change in sense

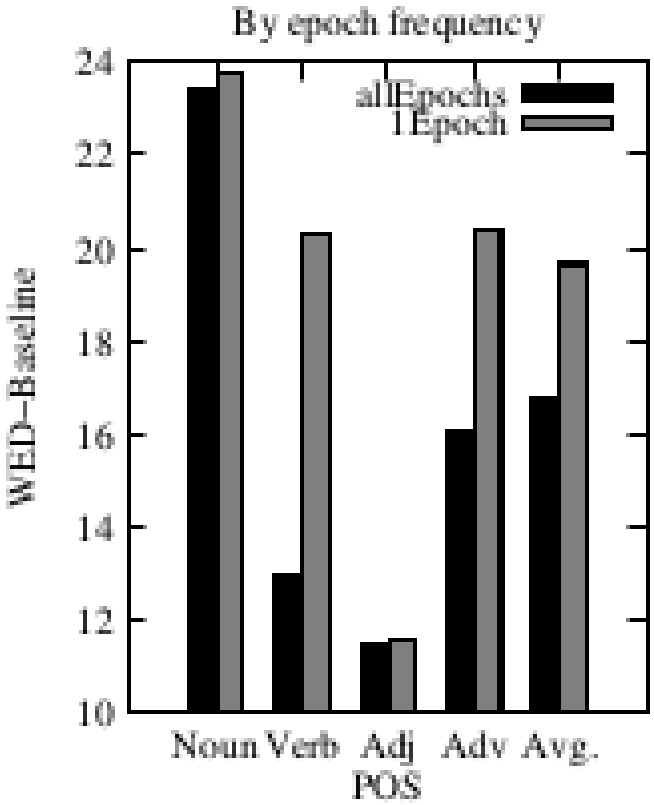
Results and Discussion contd.

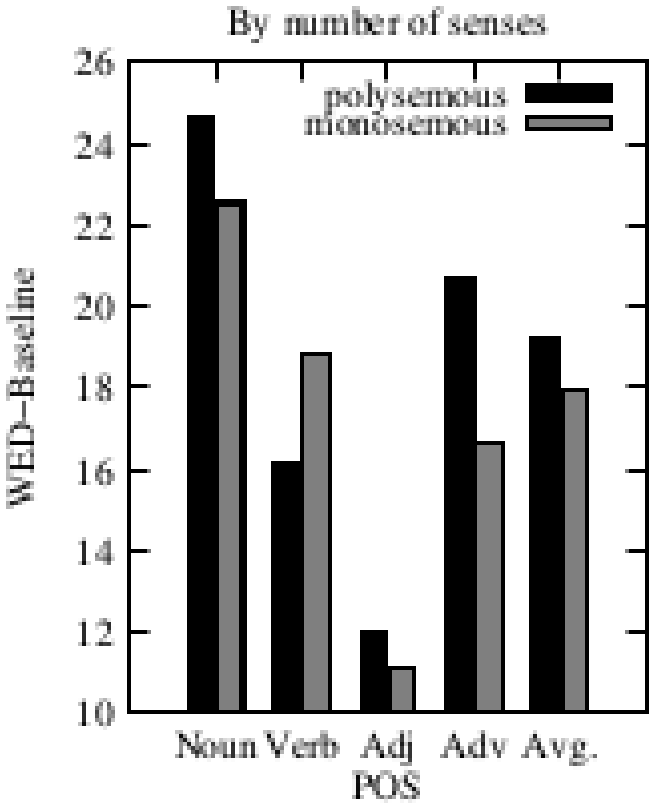
- Frequency change:
  - Words frequent in all epochs v/s only in one epoch

Results and Discussion contd.

- Frequency change:
  - Words frequent in all epochs v/s only in one epoch
- Sense change:
  - Monosemous words
    - Only changes in context helps
  - Polysemous words
    - Context change and
    - Change in meaning

Results and Discussion contd.





#

Conclusion

- Differences in occurences of words in different epochs
- Usage frequency and Word sense both contribute to these differences

Conclusion

- Differences in occurences of words in different epochs
- Usage frequency and Word sense both contribute to these differences
- Future Work
  - Feature selection
  - Representation to track sense changes