Identifying Expressions of Opinion in Context

Eric Breck, Yejin Choi, and Claire Cardie
Automatically gathering subjective content from text

- Beyond extracting facts: opinions, perspectives
- Document-level or sentence-level subjective tasks (well studied)
  - e.g. is this camera review positive or negative?
- Subjective extraction tasks (less well studied)
  - e.g. given relevant newspaper articles
    - Produce a summary: who blames who for the handling of Hurricane Katrina?
    - Answer a question: what groups are in favor of a particular law?
  - For these we need *expression-level opinion analysis*
Expression-level opinion analysis

Vedrine expressed extreme concern about the White House statement

- Who holds this opinion? (Vedrine)
- Is it positive or negative? (negative)
- How strong is it? (extremely)
- What is the target of this opinion? (the White House statement)

Central to all of these: identifying expressed extreme concern
Identifying opinion expressions (two kinds)

- Direct subjective expressions
  - Explicitly express an attitude or opinion
  - Vedrine expressed extreme concern
  - They were killed by sharpshooters faithful to the president

- Expressive subjective elements
  - Indicate subjectivity by choice of words
  - The so-called expert was wrong today.
  - Tsvangirai called the elections “highway robbery”.

- (Wiebe et al, 2005)
The task

• Given new text, identify the opinion expressions in it
  - Input: Vedrine expressed extreme concern about the White House statement.
  - Output: Vedrine *expressed extreme concern* about the White House statement.
  - Separately identify DSE and ESE (different roles)

• Approach: supervised machine learning
  - Learn a model from documents annotated by hand
  - Predict and evaluate on unseen data
The model

- Linear-chain conditional random fields (Lafferty et al, 2001)
  - Handle large feature sets
  - Model sequences
- One feature vector per word

Vedrine expressed extreme concern

<target=False, word=Vedrine, part-of-speech=Noun, ...>

<target=True, word=expressed, part-of-speech=Verb, ...>
Features

• Lexical (current word, nearby words)
• Syntactic (part-of-speech of current word, syntactic constituent type)
• Dictionary-based
  – Is the current word on a list of opinion expressions from the literature? (Wilson et al, 2005)
  – Is the current word likely a communication word? (Levin, 1993; FrameNet)
  – What are the superordinate categories of the current word in WordNet? (Miller, 1995)
    • For concern: anxiety, emotion, feeling, psych state
Baselines for comparison

- How well does a dictionary-lookup do?
- (Wiebe & Riloff 2005) (Wilson et al, 2005) – clues to subjectivity
- A clue is likely an expression of opinion
- Doesn't distinguish DSE from ESE
Data

- Multi-Perspective Question Answering set (MPQA) (Wiebe et al, 2005)
- 535 newswire documents (FBIS)
  - 135 used for development
  - Results are cross-validation on remaining 400
- Annotated for DSE and ESE (and more)
Experiment: Overall

Performance is good, close to human
Experiment: fair baseline comparison

- Baselines don't distinguish DSE from ESE
- Re-training models on (DSE union ESE), still beat baselines
Experiment: feature ablation

- Other dictionaries are helpful, but WordNet subsumes the others
Other experiments (see paper)

- Word-by-word model or sequence?
- Jointly or separately model DSEs and ESEs?
- Alternate evaluation metrics: strict boundaries or not?
Conclusions and Future work

● We can recover expressions of opinion with near-human-level accuracy

● Next: use this in opinion analysis applications
  – Opinion summarization (in progress)
  – Opinionated question-answering (future work)
**Experiment: Order 0 or Order 1?**

- **Overlap Recall & F**: Order 0 > Order 1
- **Overlap Precision & F**: Order 0 > Order 1
- **Exact Recall & F**: Order 0 > Order 1
- **Exact Precision & F**: Order 0 > Order 1
Experiment: 2way vs 3way for DSE

- **2way > 3way** for precision, not recall
Evaluation

• Standard information extraction metrics
  – Precision = fraction of predicted items which are correct
  – Recall = fraction of correct items which were predicted
  – F-measure = 2PR/(P+R)

• Problem in this domain: fuzzy boundaries
  – Vedrine expressed extreme concern about ...
  – We count this as correct (as did annotators)