Eliciting Subjectivity and Polarity Judgements on Word Senses

Fangzhong Su & Katja Markert

School of Computing
University of Leeds

August 23, 2008
Motivation I

- A popular task
  - Annotating word subjectivity or polarity: subjective/objective, or positive/negative/neutral

  “positive” $\rightarrow$ subjective; “catch” $\rightarrow$ neutral

- Existing problems
  - Subjectivity-ambiguous or polarity-ambiguous words

  (1) positive, electropositive—having a positive electric charge (objective)
  (2) plus, positive—involving advantage or good (subjective)
  (3) catch—a hidden drawback; “it sounds good but what’s the catch?” (negative)
  (4) catch, match—a person regarded as a good matrimonial prospect (positive)
Motivation II

- Human judgement difficulty in opinions
- Impact on other tasks or applications
  - Word sense disambiguation (Wiebe and Mihalcea, ACL’06)
1. Definition of Subjectivity and Polarity

2. Human Annotation Study

3. The Effect of Hierarchical Annotation

4. Annotation Bias

5. Conclusion and Future Work
Subjectivity and Polarity Property of Senses

- **Subjectivity**
  - Refer to private states: emotions, judgements, or mental states (doubts, beliefs or speculations)
  - Categories: subjective (S), objective (O), and both (B)

- **Polarity**
  - Refer to positive or negative connotations associated with a sense
  - Categories: positive (P), negative (N), varying (V), and no-polarity (NoPol)

**Difference between subjectivity and polarity**

- Subjectivity: private state
- Polarity: positive/negative connotation
Subjectivity Property of Senses

Definition
Follow Wiebe and Mihalcea (ACL’06)

- Subjective
  Refer to private states: emotions, judgements, and mental states (doubts, beliefs, and speculations)

- Objective
  Refer to persons, objects, actions or states without inherent emotion, judgement or mental states

- Both
  Conflate both opinionated and objective expressions
Examples

- angry—feeling or showing anger; “angry at the weather”; “angry customers”; “an angry silence” (Subjective—emotion)
- beautiful—aesthetically pleasing (Subjective—aesthetic assessment)
- alarm clock, alarm – a clock that wakes sleeper at preset time (Objective—non-judgemental reference to object)
- lawyer, attorney – a professional person authorized to practice law; conducts lawsuits or gives legal advice (Objective—non-judgemental reference to person)
- alcoholic, alky, dipsomaniac, boozer, lush, soaker, souse—a person who drinks alcohol to excess habitually (Both)

All examples are from WordNet 2.0
Polarity Property of Sense

Polarity of Subjective Senses
- S:P—private states that express a positive attitude, emotions or judgements
- S:N—private states that express a negative attitude, emotion or judgement
- S:V—polarity is varying by context or user

Polarity of Objective Senses
- O:P—objective sense with strong positive connotation
- S:N—objective sense with strong negative connotation
- O:NoPol—objective sense with no strong, generally shared connotations
Examples

- good, right, ripe – most suitable or right for a particular purpose; “a good time to plant tomatoes”; “the right time to act”; (S:P)
- hot – very unpleasant or even dangerous; “make it hot for him”; “in the hot seat” (S:N)
- aloof, distant, upstage—remote in manner; “stood apart with aloof dignity”; “a distant smile”; “he was upstage with strangers” (S:V)
- remedy, curative, cure – a medicine or therapy that cures disease or relieve pain (O:P)
- disease—an impairment of health or a condition of abnormal functioning (O:N)
- above—appearing earlier in the same text; “flaws in the above interpretation” (O:NoPol)
Hierarchy of all categories

Figure: Overview of the hierarchy over all categories
Annotation Study

**Dataset**
- Micro-WNOp corpus\(^2\)
- 3 Groups, 298 words with 1105 WordNet senses
- Representative of the part-of-speech distribution in WordNet

**Annotation Procedures**
- Annotators—2 near native English speakers
- Annotation Guidelines
- Annotate each item independently

---

\(^2\)http://www.unipv.it/wnop/micrownop.tgz
## Agreement Study

### Training:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>S:N</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>S:P</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>S:V</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>13</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>O:NoPol</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>51</td>
</tr>
<tr>
<td>O:N</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>O:P</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>total</td>
<td>3</td>
<td>14</td>
<td>9</td>
<td>14</td>
<td>61</td>
<td>6</td>
<td>3</td>
<td>110</td>
</tr>
</tbody>
</table>

- Agreement: 83.6%  Kappa: 0.76
- Categories with low reliability: B and S:V
Agreement Study

Testing:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>S:N</td>
<td>0</td>
<td>41</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>42</td>
</tr>
<tr>
<td>S:P</td>
<td>0</td>
<td>0</td>
<td>65</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>71</td>
</tr>
<tr>
<td>S:V</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>17</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>O:NoPol</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>253</td>
<td>5</td>
<td>8</td>
<td>284</td>
</tr>
<tr>
<td>O:N</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>41</td>
</tr>
<tr>
<td>O:P</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>13</td>
<td>20</td>
</tr>
<tr>
<td>total</td>
<td>17</td>
<td>58</td>
<td>80</td>
<td>31</td>
<td>257</td>
<td>30</td>
<td>23</td>
<td>496</td>
</tr>
</tbody>
</table>

- Agreement: 84.9%  Kappa: 0.77
- Single-category Kappa:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.80</td>
<td>0.84</td>
<td>0.86</td>
<td>0.49</td>
<td>0.56</td>
<td>0.68</td>
<td>0.59</td>
</tr>
</tbody>
</table>
Subjectivity Distinction Only

- Merging subcategories:
  S—S:V, S:P, and S:N;
  O—O:NoPol, O:P, and O:N;
  B (remain)

Results

- Agreement: 90.1%  Kappa: 0.79
- Single-category Kappa:

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>O</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.82</td>
<td>0.80</td>
<td>0.49</td>
</tr>
</tbody>
</table>
Polarity Distinction Only

- Merging subcategories:
  - N—O:N and S:N;
  - P—O:P and S:P;
  - B (remain);
  - V—S:V;
  - NoPol—O:NoPol

Results

- Agreement: 89.1%  Kappa: 0.83
- Single-category Kappa:

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>P</th>
<th>B</th>
<th>V</th>
<th>NoPol</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.92</td>
<td>0.85</td>
<td>0.49</td>
<td>0.56</td>
<td>0.86</td>
</tr>
</tbody>
</table>
Annotation Bias I

- **Individual perspective or bias**

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>N</th>
<th>P</th>
<th>V</th>
<th>NoPol</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>N</td>
<td>0</td>
<td>80</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>83</td>
</tr>
<tr>
<td>P</td>
<td>1</td>
<td>0</td>
<td>85</td>
<td>4</td>
<td>1</td>
<td>91</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>17</td>
<td>3</td>
<td>27</td>
</tr>
<tr>
<td>NoPol</td>
<td>9</td>
<td>6</td>
<td>10</td>
<td>6</td>
<td>253</td>
<td>284</td>
</tr>
<tr>
<td>total</td>
<td>17</td>
<td>88</td>
<td>103</td>
<td>31</td>
<td>257</td>
<td>496</td>
</tr>
</tbody>
</table>

- **Conflation of near-synonym terms which differ in sentiment property**

  (1)alcoholic, alky, dipsomaniac, boozer, lush, soaker, souse—a person who drinks alcohol to excess habitually
**Annotation Bias II**

- Connotation bias in a gloss or its hierarchical organization

(2) Iran, Islamic Republic of Iran, Persia—a theocratic islamic republic in the Middle East in western Asia; Iran was the core of the ancient empire that was known as Persia until 1935; rich in oil; involved in state-sponsored terrorism

(3) skinhead—a young person who belongs to a British or American group that shave their heads and gather at rock concerts or engage in white supremacist demonstrations

skinhead ← bully, tough, hooligan, ruffian, roughneck, rowdy, yob, yobo, yobbo—(a cruel and brutal fellow)
Subjectivity-ambiguous words: 32.5% (97/298)

Polarity-ambiguous words:
- 3.4% (10/298) of words have at least one positive and one negative polarity
- With further 14.8% (44/298) of words having varying (S:V) polarity
Conclusion and Future Work

**Conclusion**
- Difference between subjectivity and polarity
- A substantial proportion of words are subjectivity-ambiguous (polarity-ambiguous)
- Hierarchical annotation affects human agreement significantly
- Annotation bias

**Future Work**
- Refine guidelines for the more difficult categories
- Perform larger-scale annotation with more annotators
- Use the annotated dataset to explore learning algorithms for the automatic detection of subjectivity and polarity properties of word sense
Any questions?