Overview: Opinion Mining

This *phone* was a *bad decision*.

I am *not a fan* of *vintage themes*.

*Nikon cameras do wonder with colors!*
Overview: Subjunctive Opinion Mining

- I wish this phone came with a colored body.
- I prefer modern themes in hotels.
- I suggest to keep a set of normal, macro and telephoto lens.
Outline

1. Context
2. Related Work
3. Proposition
4. Current Status
5. Future Objectives
Context

1. Opinion = Information
2. Opinion Explosion -> Automatic analysis
Context: Opinion Mining and Sentiment Analysis

The image shows a Twitter sentiment analysis of US Election Candidates' tweets on October 22 from 5-9 pm. The analysis is divided by city, with a map indicating sentiment by city for both Obama and Romney. The overall sentiment is also calculated.

Overall Twitter Sentiment by Candidate:

- Obama: 57.92
- Romney: 58.10

Sentiment by City:

<table>
<thead>
<tr>
<th>Candidates</th>
<th>Columbia</th>
<th>Columbus</th>
<th>Concord</th>
<th>Denver</th>
<th>Des Moines</th>
<th>Madison</th>
<th>Raleigh</th>
<th>Tallahassee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obama</td>
<td>59.50</td>
<td>56.40</td>
<td>61.20</td>
<td>61.20</td>
<td>54.10</td>
<td>68.40</td>
<td>56.60</td>
<td>57.80</td>
</tr>
<tr>
<td>Romney</td>
<td>60.40</td>
<td>59.40</td>
<td>60.00</td>
<td>56.10</td>
<td>57.20</td>
<td>55.30</td>
<td>59.10</td>
<td>59.10</td>
</tr>
</tbody>
</table>
Context: Opinion Mining and Sentiment Analysis

Twitter Sentiment
Type in a word and we'll highlight the good and the bad

iphone
Search Save this search

Sentiment analysis for iphone

Sentiment by Percent
Negative (44%)
Positive (56%)

Sentiment by Count
Positive (1996252)
Negative (1568474)
1. Two terms often used inter-changeably.

2. Definition: automatic analysis of opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes.
**Context:** Sentiment Opinions

<table>
<thead>
<tr>
<th>Sentences</th>
<th>Sentiment Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>I love my new iPad.</td>
<td>positive</td>
</tr>
<tr>
<td>I did a big mistake by buying this product.</td>
<td>negative</td>
</tr>
<tr>
<td>I bought this phone 2 months back.</td>
<td>neutral</td>
</tr>
</tbody>
</table>
**Context:** Other Opinions?

<table>
<thead>
<tr>
<th>Sentences</th>
<th>Possible Opinion Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>If only it played stand alone avi files</td>
<td>wish, suggestion</td>
</tr>
<tr>
<td>After using it for a week, I wish I had purchased the ipod</td>
<td>wish</td>
</tr>
<tr>
<td>Symantec should give us all a refund and apology</td>
<td>demand</td>
</tr>
</tbody>
</table>
Related Work

Wish Mining (Goldberg et al, 2009):
- Identification of Wish sentences in opinion text.
- Features: words, automatically extracted patterns from a corpus of new year wishes.

Dataset:
- Manual wish/non-wish annotations for English language.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Accuracy</th>
<th>Total Sentences</th>
<th>% wishes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Political discussion</td>
<td>0.80 (+/-) 0.03</td>
<td>6379</td>
<td>34%</td>
</tr>
<tr>
<td>Product reviews</td>
<td>0.56 (+/-) 0.16</td>
<td>1235</td>
<td>12%</td>
</tr>
</tbody>
</table>
Related Work

Wish Identification (Ramanand et al, 2010):
- Identification of ‘suggestions’ and ‘buy’ wishes only.
- Manually formulated rules.

Dataset:
- Product reviews.
- Not available.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Precision</th>
<th>Recall</th>
<th>Total sentences</th>
<th>suggestion wishes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Banking</td>
<td>0.30</td>
<td>0.60</td>
<td>15408</td>
<td>5%</td>
</tr>
<tr>
<td>Product 1</td>
<td>0.80</td>
<td>0.53</td>
<td>4240</td>
<td>1%</td>
</tr>
<tr>
<td>iPod</td>
<td>0.588</td>
<td>0.50</td>
<td>21147</td>
<td>1%</td>
</tr>
<tr>
<td>Product reviews (Goldberg et al, 2009)</td>
<td>0.57</td>
<td>0.39</td>
<td>1236</td>
<td>3%</td>
</tr>
</tbody>
</table>
Related Work

Others:

- Sentiment Analysis: improved accuracy by removing ‘suggestion’ verbs from the verb features (Chesley et al, 2006)

- Recommendation Systems: suggestion extraction for product features, recommend a product with the desirable feature.

Works named according to the applications – wish mining, suggestion extraction, sentiment analysis.
Related Work

Others:

- Sentiment Analysis: improved accuracy by removing ‘suggestion’ verbs from the verb features (Chesley et al, 2006)

- Recommendation Systems: suggestion extraction for product features, recommend a product with the desirable feature.

Works named according to the applications – wish mining, suggestion extraction...
<table>
<thead>
<tr>
<th>Examples</th>
<th>Possible Opinion Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>If only it played stand alone avi files.</td>
<td>wish, suggestion</td>
</tr>
<tr>
<td>After using it for a week, I wish I had purchased the ipod.</td>
<td>wish</td>
</tr>
<tr>
<td>Symantec should give us all a refund and apology.</td>
<td>demand</td>
</tr>
</tbody>
</table>

Proposed term: Subjunctive Opinions
Proposition

Subjunctive Mood:

Very common language phenomenon in Indo-European languages, which is a verb mood typically used in subordinate clauses to express a wish, emotion, possibility, judgment, opinion, necessity, or action that has not yet occurred.

Extensively researched in linguistic community.
Proposition

Proposed Definition:

Subjunctive Opinions (SO) speak about various states of unreality like wish, suggestion, advice, request, possibility, necessity, or action that has not yet occurred.

SO are challenges/outliers for most Sentiment Analysis techniques. Can contain implicit positive/negative sentiments. Eg. “Symantec should give us all a refund and apology” is negative.
Current Status

Objective:
Classification of opinions into Subjunctive and non-Subjunctive.

Approaches:
- Statistical Classification: Features identified using the existing rules for Subjunctive Mood.
- Rule Based Classification: Automatic subjunctive pattern extraction

Datasets:
- Available datasets for wish mining.
- Our datasets
  - Example sentences of Subjunctive Mood.
  - Manual annotations of opinions from social media.
Current Status : Experiment 1

Approach:

- Statistical Classification.
- Features:
  
  Formulae of Subjunctive Mood (Guan, 2012) = **Parameters** + **Variables**

Example:

“He could have taken a taxi”

Formula = ‘**could have**’(N,S-Z)

N, A, V, Z, S = noun, adjective, verb, subject and sentence respectively.
Current Status : Experiment 1

Classifier:
Support Vector Machine

Features:
1. Condition indicator: “if”
2. Modal Verbs: could, should, would etc
3. Verbs: expansion of categories – wish, advice, want, urge, require; using VerbNet; total 28
4. Adjectives: expansion of necessary, urgent, desirable; using WordNet; total 9
5. Number of clauses: I suggest, you better buy an iPod = 2 clauses
6. Unigrams
# Current Status: Experiment 1

## Data:

<table>
<thead>
<tr>
<th>id</th>
<th>Data type</th>
<th>Intended purpose</th>
<th>Size (in Sentences)</th>
<th>% subjunctive opinions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Political discussion</td>
<td>Training</td>
<td>6379</td>
<td>34%</td>
</tr>
<tr>
<td>2</td>
<td>Product reviews</td>
<td>Training</td>
<td>1235</td>
<td>12%</td>
</tr>
<tr>
<td>3</td>
<td>Subjunctive Mood Examples (new)</td>
<td>Training</td>
<td>229</td>
<td>100%</td>
</tr>
<tr>
<td>4</td>
<td>Book Reviews (new)</td>
<td>Evaluation</td>
<td>604</td>
<td>22%</td>
</tr>
</tbody>
</table>
Current Status : Experiment 1

Training Datasets:

- Dataset 1: political discussion (#1)
- Dataset 2: product reviews (#2)
- Dataset 3: Dataset 1+ Dataset 2
- Dataset 4: All Training (#1,2,3)
## Current Status: Experiment 1

### Results: (highest Classification Accuracy)

<table>
<thead>
<tr>
<th>Training Dataset</th>
<th>SO Precision</th>
<th>SO Recall</th>
<th>Classification Accuracy</th>
<th>Best Accuracy Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset 1</td>
<td>.69</td>
<td>.45</td>
<td>.81</td>
<td>all</td>
</tr>
<tr>
<td>Dataset 2</td>
<td>.50</td>
<td>.02</td>
<td>.75</td>
<td>all</td>
</tr>
<tr>
<td>Dataset 3</td>
<td>.86</td>
<td>.17</td>
<td>.78</td>
<td>unigrams</td>
</tr>
<tr>
<td>Dataset 4</td>
<td>.70</td>
<td>.75</td>
<td>.859</td>
<td>All except unigrams</td>
</tr>
</tbody>
</table>
Future Objectives

Short term:

- Improve current results for subjunctive opinion identification.
- Perform rule based classification.
- Expansion of new dataset.

Long Term:

- Suggestion extraction from subjunctive opinions.
- Suggestion extraction and summarisation.
Feedback?