Fine-Grained Sentiment Analysis with Structural Features

Cacilia Zirn, Mathias Niepert, Heiner Stuckenschmidt, Michael Strube
Motivation

Goal:

- **Sentiment analysis:**
  - determining the polarity of a text with respect to a particular topic.

Existing Solutions:

- Document level
- Sentence level
Motivation

Problem:

- Different opinions can even be uttered in the same sentence.

E.g.

Despite the pretty design(Positive) I wound never recommend it(Negative), because the sound quality is unacceptable(Negative).
Motivation

A New Fine-Grained solution:

- Sentiment analysis on subsentence level.
- Motivated by Rhetorical Structure Theory (Mann and Thompson, 1988), every text consists of elementary segments that are connected by relations.

e.g.
- s1 = Despite the pretty design
- s2 = I wound never recommend it
- s3 = because the sound quality is unacceptable

s1 -(concession)- s2 -(cause-explanation-evidence)- s3
Method

Basic idea:

• Integration the heterogeneous features such as polarity scores from sentiment lexicons and neighborhood relations between segments.

• Employ Markov logic networks (MLN) as the framework for combing numerical and structure features.
Method

Markov Networks:

- A Markov network $M$ is an undirected graph whose nodes represent a set of random variables $X = \{X_1, \ldots, X_n\}$ and whose edges model direct probabilistic interactions between adjacent nodes. More formally, a distribution $P$ is a log-linear model over a Markov network $M$ if it is associated with:
  
  - a set of features $\{f_1(D_1), \ldots, f_k(D_k)\}$, where each $D_i$ is a clique in $M$ and each $f_i$ is a function from $D_i$ to $\mathbb{R}$.
  
  - a set of real-valued weights $w_1, \ldots, w_k$, such that
    \[
    P(X = x) = \frac{1}{Z} \exp \left( \sum_{i=1}^{k} w_i f_i(D_i) \right),
    \]
    where $Z$ is a normalization constant.
Method

Markov Logic Networks[1,2]:

- A Markov Logic Network (MLN) $L$ is a set of pairs $(F_i, w_i)$ where:
  - $F_i$ is a formula in first-order logic
  - $w_i$ is a real number (the weight of the formula)

- Applied to a finite set of constants $C = \{c_1, \ldots, c_{|C|}\}$ it defines a Markov network $M_{L,C}$:
  - $M_{L,C}$ has one binary node for each possible grounding of each atom in $L$. The value of the node is 1 if the ground atom is true, 0 otherwise.
  - $M_{L,C}$ has one feature for each possible grounding of each formula $F_i$ in $L$. The value of the feature is 1 if the ground formula is true, 0 otherwise. The weight of the feature is the weight $w_i$ of the corresponding formula.
Method

Markov Logic Networks:

- A ground MLN specifies a joint probability distribution over possible worlds (i.e. truth value assignments to all ground atoms)
- The probability of a possible world $x$ is:

$$p(x) = \frac{1}{Z} \exp \left( \sum_{i=1}^{F} w_{i} n_{i}(x) \right)$$

where:

- the sum ranges over formulas in the MLN (i.e. clique templates in the Markov Network)
- $n_{i}(x)$ is the number of true groundings of formula $F_{i}$ in $x$
Method

Markov Logic Networks:

- A MLN with two (weighted) formulas:
  \[ w_1 \quad \forall x \ (\text{Bird}(x) \Rightarrow \text{Flies}(x)) \]
  \[ w_2 \quad \forall x, y \ (\text{Predates}(x, y) \land \text{Bird}(y) \Rightarrow \text{Bird}(x)) \]

- applied to a set of two constants \{\text{Sparrow}, \text{Eagle}\}
- generates the Markov Network shown in figure
Method

Markov Logic Networks:

\[ p(x) = \frac{1}{Z} \exp(2w_1 + w_2) \]

\[ p(x) = \frac{1}{Z} \exp(2w_1 + 4w_2) \]
Method

Markov Logic Networks:

- Inference:
  1. (conditional) probability inference
  2. Maximum a-posteriori inference

\[
\arg\max_y \sum_{i=1}^k w_i f_i(D_i).
\]

- Parameter Learning:
  - Employing the Voted perceptron learner for the experiments
Method

Markov Logic Formulation:

• Polarity formulas:
  • A segment is positive or negative but cannot be positive and negative at the same time:
    \[ \forall x : \neg \text{positive}(x) \implies \text{negative}(x) \]
    \[ \forall x : \text{negative}(x) \implies \neg \text{positive}(x) \]

  • Numerical a-priori features such as the polarity scores of individual segments provided by external lexical resources
    \[ \forall x : \text{positive-source}_\ell(x) \iff \text{positive}(x) \]
    \[ \forall x : \text{negative-source}_\ell(x) \iff \text{negative}(x) \]
Method

Markov Logic Formulation:

• Neighborhood Relations:
  • Intuition: neighboring segments are more likely to have the same polarity.
  • $pre$ relation: $pre(x,y)$ if segment $x$ precedes segment $y$.

\[
\forall x, y : pre(x, y) \land positive(x) \Rightarrow positive(y) \\
\forall x, y : pre(x, y) \land negative(x) \Rightarrow negative(y)
\]

• The weights of above formulas are learned during training
Method

Markov Logic Formulation:

• Neighborhood Relations:
  • Intuition: neighboring segments are more likely to have the same polarity.
  • pre relation: \( \text{pre}(x,y) \) if segment \( x \) precedes segment \( y \).

\[
\forall x, y : \text{pre}(x, y) \land \text{positive}(x) \Rightarrow \text{positive}(y)
\]
\[
\forall x, y : \text{pre}(x, y) \land \text{negative}(x) \Rightarrow \text{negative}(y)
\]

• The weights of above formulas are learned during training
Method

Markov Logic Formulation:

- **Discourse Relations:**
  - Intuition: employ discourse relations between discourse segments
  - *contrast and ncontrast* relations:
    \[
    \begin{align*}
    \forall x, y : \text{contrast}(x, y) \land \text{positive}(x) & \Rightarrow \text{negative}(y) \\
    \forall x, y : \text{contrast}(x, y) \land \text{negative}(x) & \Rightarrow \text{positive}(y) \\
    \forall x, y : \text{ncontrast}(x, y) \land \text{positive}(x) & \Rightarrow \text{positive}(y) \\
    \forall x, y : \text{ncontrast}(x, y) \land \text{negative}(x) & \Rightarrow \text{negative}(y)
    \end{align*}
    \]
  - Again, the weights of above formulas are learned during training
Experiment

Data:

• A subset of the Multi-Domain Sentiment Dataset
  
  • Step 1: Pick three categories:
    • Cell Phones & Service
    • Gourmet Food
    • Kitchen & Housewares

  • Each category consists of up to 100 reviews, each review is classified as positive/negative according the rating stars.

  • Step 2: Pick 20 longest positive and 20 longest negative for each (amount 120 reviews).

<table>
<thead>
<tr>
<th>Topic</th>
<th>p</th>
<th>n</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Phones &amp; Service</td>
<td>1392</td>
<td>1785</td>
<td>3177</td>
</tr>
<tr>
<td>Gourmet Food</td>
<td>990</td>
<td>616</td>
<td>1606</td>
</tr>
<tr>
<td>Kitchen &amp; Housewares</td>
<td>1188</td>
<td>1405</td>
<td>2593</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td><strong>3570</strong></td>
<td><strong>3806</strong></td>
<td><strong>7376</strong></td>
</tr>
</tbody>
</table>

Table 1: Amount of positive (p) and negative (n) segments.
Experiment

Data:

- Gold Standard
  - Three annotators

- Label: positive negative or neutral for each passage.
  - Passage: a sequence of words sharing the same opinion
  - Agreement:
    - $k = 0.40$ to $0.45$ for negative reviews (Fail)
    - $k = 0.60$ to $0.84$ for positive reviews (Strong agreement)

- Polarity label for discourse segment:
  - Majority voting of all tokens belonging to the segment
Experiment

Details:

• Polarity features:
  • SentiWordNet
  • Taboada and Grieve's Turney Adjective List (TGL)
  • Unigram Lexicon (UL)

• Discourse Parsing
  • Employ the discourse parser HILDA (Verle & Prednginger, 2009)
  • Two tasks:
    • Splitting the review text into discourse segments
    • Determining the discourse relations between segments
    • contrast v.s. all others (ncontrast)

• 10 fold cross-validation
Experiment

Results:

<table>
<thead>
<tr>
<th></th>
<th>positive</th>
<th></th>
<th></th>
<th></th>
<th>negative</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F</td>
<td>A</td>
<td>P</td>
<td>R</td>
<td>F</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>majority baseline</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td>51.60</td>
<td>100.00</td>
<td>68.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM</td>
<td>57.05</td>
<td>43.06</td>
<td>49.08</td>
<td></td>
<td>56.44</td>
<td>69.47</td>
<td>62.28</td>
<td>56.66</td>
<td></td>
</tr>
<tr>
<td>MLN_polarity</td>
<td>53.21</td>
<td>69.58</td>
<td>60.31</td>
<td></td>
<td>59.90</td>
<td>42.62</td>
<td>49.80</td>
<td>55.67</td>
<td></td>
</tr>
<tr>
<td>MLN_neighborhood</td>
<td>66.38</td>
<td>72.94</td>
<td><strong>69.50</strong></td>
<td></td>
<td>72.02</td>
<td>65.34</td>
<td><strong>68.52</strong></td>
<td><strong>69.02</strong></td>
<td></td>
</tr>
<tr>
<td>MLN_contrast</td>
<td>61.39</td>
<td>73.47</td>
<td>66.89</td>
<td></td>
<td>69.48</td>
<td>56.65</td>
<td>62.41</td>
<td>64.79</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Results (%) for the different systems. P = precision, R = recall, F = F-measure, A = accuracy

Figure 1: Accuracy values of the various algorithms for the 10 different cross-validation folds.
References
