Expanding Domain Sentiment Lexicon through Double Propagation

Presenter: He Jiang

Report based on the following materials:

2) Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. Expanding Domain Sentiment Lexicon through Double Propagation. IJCAI’09
Outline

- Part 1: Opinion Mining and Sentiment Analysis
- Part 2: Expanding Domain Sentiment Lexicon through Double Propagation
Part 1: Opinion Mining and Sentiment Analysis

- Opinion on the Web
- Applications
- Opinion Search
- Related Problems
Part 1: Opinion Mining and Sentiment Analysis

Opinion on the Web

- Two main types of information on the web
  - facts
  - opinion

- Where’s opinion
  - review sites
  - forums
  - discussion groups
  - blogs
Part 1: Opinion Mining and Sentiment Analysis

Opinion on the Web

Two types of opinions

- Direct Opinions: sentiment expressions on some objects, e.g. products, events, topics, persons
  - e.g. “the picture quality of this camera is great”

- Comparisons: relations expressing similarities or differences of more than one object
  - e.g. “car x is cheaper than car y”
Part 1: Opinion Mining and Sentiment Analysis

Applications

Businesses and organizations: product and service benchmarking.
Market intelligence.
- Business spends a huge amount of money to find consumer sentiments and opinions.
- Consultants, surveys and focused groups, etc.

Individuals: interested in other’s opinions when
- Purchasing a product or using a service,
- Finding opinions on political topics,

Ads placements: Placing ads in the user-generated content
- Place an ad when one praises a product.
- Place an ad from a competitor if one criticizes a product.

Opinion retrieval/search: providing general search for opinions.
Part 1: Opinion Mining and Sentiment Analysis

Opinion Search

- Find the opinion of a person or organization (opinion holder) on a particular object or a feature of the object.
  - e.g., what is Bill Clinton’s opinion on abortion?
- Find positive and/or negative opinions on a particular object (or some features of the object), e.g.,
  - customer opinions on a digital camera.
  - public opinions on a political topic.
- Find how opinions on an object change over time.
- How object A compares with Object B?
  - Gmail vs. Hotmail
Part 1: Opinion Mining and Sentiment Analysis

- **Related Problems**
  - **Sentiment classification**
    - classify words' sentiment expressed by authors (positive, negative, neutral)
  - **Feature-based opinion extraction and summarization**
    - A product consists of a few features (components), what's the opinion over each feature
  - **Comparative sentence and relation extraction**
  - **Visual Summarization & Comparison**
Part 1: Opinion Mining and Sentiment Analysis

- Related Problems
  - Visual Summarization & Comparison
    - Summary of reviews of Digital camera 1
    - Comparison of reviews of Digital camera 1 and Digital camera 2

The University of Vermont
Part 2: Expanding Domain Sentiment Lexicon through Double Propagation

- Motivation
- Related work
- Basic Idea
- Sentiment Word Extraction
- Polarity Assignment
- Experiments and Discussions
Part 2: Expanding Domain Sentiment Lexicon through Double Propagation

**Motivation**

Sentiment words: words that convey positive or negative sentiment polarities (attitude)

A comprehensive sentiment lexicon (dictionary) is essential
- opinion expressions vary significantly among different domains, i.e., domain dependent
- no universal sentiment lexicon to cover all domains

How to extract domain-dependent sentiment words based on a set of seeding sentiment words?
Part 2: Expanding Domain Sentiment Lexicon through Double Propagation

- Related Work
  - Sentiment Analysis can be conducted at
    - word (like this paper)
    - expression
    - sentence
  - Sentiment Word Analysis on Word Level
    - Corpora-based approaches (this paper falls in this category)
      - Kanayama and Nasukawa 2006
        - Extract domain specific sentiment words in Japanese text
        - Idea: 1) They exploit sentiment coherency within sentence and among sentences to extract sentiment candidates; 2) then use a statistical method to determine whether a candidate is correct or not.
        - Key difference with double propagation: Double propagation exploits the relationships between sentiment words and product features in extraction process.
  - dictionary-based approaches
Part 2: Expanding Domain Sentiment Lexicon through Double Propagation

Basic Idea

- **Input:** a set of sentiment words
- **Output:** a set of sentiment words & a set of features

```
input: a set of sentiment words
output: a set of sentiment words & a set of features
```

```
sentiment words -> new sentiment words
sentiment words -> new features
features -> new sentiment words
features -> new features
```
Part 2: Expanding Domain Sentiment Lexicon through Double Propagation

- **Sentiment Word Extraction**
  - 4 tasks:
    - Sentiment words $\rightarrow$ new sentiment words
    - Sentiment words $\rightarrow$ new features
    - Features $\rightarrow$ new sentiment words
    - Features $\rightarrow$ new features

- **Model:**
  - (Extraction Rules based on Relations)
  - Dependency parser Minipar
        
  - Sentiment words
  - Features
Part 2: Expanding Domain Sentiment Lexicon through Double Propagation

- **Sentiment Word Extraction**

  - Relations of sentiment words and features

  ![Diagram](image)

  **Fig. 1.** Different relations between words $A$ and $B$. (a) and (b) are two direct relations; (c) and (d) are two indirect relations.

- e.g. “I love iPod”

  - both “I” and “iPod” depend on the verb “love” with the relations of subj and obj respectively. Here, subj: “I” is the subject of “love”, while obj means that “iPod” is the object of “love”.

The University of Vermont
Part 2: Expanding Domain Sentiment Lexicon through Double Propagation

Sentiment Word Extraction

Extraction Rules based on Relations

POS: part-of-speech
s, f: word/feature to be extracted
{S}, {F}: known sentiment words/features
{JJ}: adjectives and their variants
e.g. smart, smarter, smartest
{NN}: nouns
e.g. picture, pictures
{CONJ}: conjunction
e.g. and, or
{MR}: dependency relations between sentiment words and features, such as mod, subj, obj, pnmod

Table 1. Rules for sentiment word and feature extraction. Column 2 is the observed relations between two words, column 3 shows the constraints on the observed relations and the final column is the result. The arrows mean dependency. For example, $S \rightarrow S\text{-Dep} \rightarrow F$ means $S$ depends on $F$ through a relation of $S\text{-Dep}$. 

<table>
<thead>
<tr>
<th>R1</th>
<th>Observations</th>
<th>Constraints</th>
<th>Outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>$S_{j0} \rightarrow S_{j0}\text{-Dep} \rightarrow S_{j0}$</td>
<td>$S_{j0} \in {S}$, $S_{j0}\text{-Dep} \in {\text{CONJ}}$, $\text{POS}(S_{j0}) \in {\text{JJ}}$</td>
<td>$s = S_{j0}$</td>
</tr>
<tr>
<td>R2</td>
<td>$S \rightarrow S\text{-Dep} \rightarrow H \leftarrow S\text{-Dep} \leftarrow S_j$</td>
<td>$S_j \in {S}$, $S\text{-Dep} = S_j\text{-Dep}$, $\text{POS}(S_j) \in {\text{JJ}}$</td>
<td>$s = S_j$</td>
</tr>
<tr>
<td>R3</td>
<td>$S \rightarrow S\text{-Dep} \rightarrow F$</td>
<td>$F \in {F}$, $S\text{-Dep} \in {\text{MR}}$, $\text{POS}(S) \in {\text{JJ}}$</td>
<td>$s = S$</td>
</tr>
<tr>
<td>R4</td>
<td>$F_{j0} \rightarrow F_{j0}\text{-Dep} \rightarrow F_{j0}$</td>
<td>$F_{j0} \in {F}$, $F_{j0}\text{-Dep} \in {\text{CONJ}}$, $\text{POS}(F_{j0}) \in {\text{NN}}$</td>
<td>$f = F_{j0}$</td>
</tr>
<tr>
<td>R5</td>
<td>$F \rightarrow F\text{-Dep} \rightarrow H \leftarrow F\text{-Dep} \leftarrow F_j$</td>
<td>$F_j \in {F}$, $F\text{-Dep} = F_j\text{-Dep}$, $\text{POS}(F_j) \in {\text{NN}}$</td>
<td>$f = F_j$</td>
</tr>
</tbody>
</table>
Part 2: Expanding Domain Sentiment Lexicon through Double Propagation

- Sentiment Word Extraction
  - a detailed model

Diagram:
- Corpus to sentiment words
- Stanford POS Tagger to tokens
- Dependency parser Minipar to rules
- Rules to features
- Features to sentiment words
Part 2: Expanding Domain Sentiment Lexicon through Double Propagation

- **Polarity Assignment**
  - **Observation 1**
    - Same polarity for same feature in a review
  - **Observation 2**
    - Same polarity for same sentiment word in a domain corpus
Part 2: Expanding Domain Sentiment Lexicon through Double Propagation

- **Polarity Assignment**
  - **Rules:**
    - **Rule 1: Heterogeneous rule**
      - For sentiment words extracted by known feature, and features extracted by known sentiment words, assign them the same polarities as the known ones.
    - **Rule 2: Homogeneous rule**
      - For sentiment words extracted by known sentiment words, and features extracted by known features, also assign them the same polarities as the known ones.
    - **Rule 3: Intra-review rule**
      - For new sentiment words extracted by features which are extracted in other reviews, those sentiment words cannot be determined (by Rule1). If those sentiments words only appear in this review, Rule2 cannot be applied either. Those sentiment words are set to the same polarity of the review.
      - The polarity of a review is determined by the sum of the polarities of words in this review (+1 for positive, -1 for negative, 0 for neutral).
        - If sum > 0, positive for the review
        - If sum < 0, negative
    - **Conflict resolution:**
      - For sentiment words with distinct polarities by calculation, use sum of those polarities to determine the final polarity.
Part 2: Expanding Domain Sentiment Lexicon through Double Propagation

Experiments and Discussions

Experiment Set Up

- Data set: 5 review data sets (2 for digital cameras, 1 for DVP player, 1 for MP3, 1 for cell phone)
- On average, each data set consists of 789 sentences and 63 reviews.

- 4 algorithms:
  - CRF (conditional random fields): a supervised algorithm (2001)
  - KN06
  - Prop-dep
  - noProp-dep (non-propagation version of the new approach)

- An initial positive and negative sentiment lists (654 and 1098 words, respectively)
- One data set is used as the training set for CRF
- words appearing in both 10% (20%, 50, 80%) of initial lists and the chosen training data set are feed into KN06, Prop-dep, noProp-dep
Part 2: Expanding Domain Sentiment Lexicon through Double Propagation

- Experiments and Discussions
  - Results of sentiment word extraction

Fig. 2. Precisions of CRF, KN06, noProp-dep and Prop-dep
Part 2: Expanding Domain Sentiment Lexicon through Double Propagation

- Experiments and Discussions
  - Results of sentiment word extraction

**Fig. 3.** Recalls of Init, CRF, KN06, noProp-dep and Prop-dep
Part 2: Expanding Domain Sentiment Lexicon through Double Propagation

- Experiments and Discussions
  - Results of sentiment word extraction

Fig. 4. F-scores of CRF, KN06, noProp-dep and Prop-dep
Part 2: Expanding Domain Sentiment Lexicon through Double Propagation

- Experiments and Discussions
  - Results of polarity assignment

**Fig. 5.** Average polarity assignment accuracy on correct new sentiment words
Conclusions

- Opinion mining is a promising topic in data mining, there’re a lot of new problems in this area!
- Sentiment word detection could be a starting point for opinion mining