Sentiment Analysis: A discovery challenge

Bing Liu
University of Illinois at Chicago
liub@cs.uic.edu
Introduction

- **Opinion mining or sentiment analysis**
  - computational study of opinions, sentiments, appraisal, and emotions expressed in text.
  - Reviews, Twitter, blogs, discussions, comments, etc

- **Why is it important?**
  - Opinions are key influencers of our behaviors.
  - Our beliefs and perceptions of reality are conditioned on how others see the world.
  - Whenever we need to make a decision we often seek out the opinions of others.
  - True for individuals and organizations
A Fascinating Problem!

- Intellectually challenging
  - A popular research topic in NLP, text mining, and even management sciences!
  - Although there has been so much research,
    - the progress has not been fast!

- Wide spread applications in every domain
  - More than 60 companies in USA alone
    - Many have died and many new ones are still coming
  - One CEO said “Our sentiment analysis is as bad as everyone else’s”
Abstraction (1): what is an opinion?

- Structure the unstructured

- **Id: Abc123 on 5-1-2008** “I bought an iPhone today. It is such a nice phone. The touch screen is cool. The voice quality is clear too. It is much better than my old Blackberry, which was a terrible phone and so difficult to type with its tiny keys. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, …”

- **We see:** Each opinion has a
  - target
  - **Sentiment:** positive and negative
  - opinion holder: person who holds the opinions
  - time when the opinion was given
What is an opinion?

(Hu and Liu, 2004; Liu. in NLP handbook)

- An opinion is a quintuple
  \[(e_j, a_{jk}, s_{ijkl}, h_i, t_i),\]
  where
  - \(e_j\) is a target entity.
  - \(a_{jk}\) is a aspect of the entity \(e_j\).
  - \(s_{ijkl}\) is the sentiment value of the opinion. \(s_{ijkl}\) is +ve, -ve, or neu, or a more granular rating.
  - \(h_i\) is an opinion holder.
  - \(t_i\) is the time when the opinion is expressed.

- Note the simplification: \(target = (e_j, a_{jk})\)
Structure the unstructured

- **Objective**: Given an opinionated document,
  - Discover all quintuples \((e_j, a_k, s_{ijkl}, h_i, t_l)\),
  - Or, solve some simpler forms of the problem
    - E.g., sentiment classification at the document or sentence level.

- **With the quintuples,**
  - **Unstructured Text → Structured Data**
    - Traditional data and visualization tools can be used to slice, dice and visualize the results.
    - Enable qualitative and quantitative analysis.
We need quantitative summary

"I bought an iPhone a few days ago. It is such a nice phone. The touch screen is really cool. The voice quality is clear too. It is much better than my old Blackberry, which was a terrible phone and so difficult to type with its tiny keys. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, ..."

Aspect-Based Summary:
Opinion summary on iPhone

Feature1: Touch screen
Positive: 212
- The touch screen was really cool.
- The touch screen was so easy to use and can do amazing things.

Negative: 6
- The screen is easily scratched.
- I have a lot of difficulty in removing finger marks from the touch screen.

Feature2: Voice quality

Note: We omit opinion holders
Opinion observer - visualization (Liu et al. 05)

- Summary of reviews of
  - Cell Phone 1

- Comparison of reviews of
  - Cell Phone 1
  - Cell Phone 2
Feature/aspect-based opinion summary

The HP LaserJet 1020 Printer, an excellent laser printer for the cost-conscious consumer, offers high-quality LaserJet printing in a compact size, and at a price you can afford. It is well-reviewed by users, with a high percentage of positive comments. The speed is rated at 96%, and users love the quality as good as any laserjet printer they've used. The transaction is quick and fast, and it's small and fast and very reliable.
Google Product Search (Blair-Goldensohn et al 2008)

Sony Cyber-shot DSC-W370 14.1 MP Digital Camera (Silver)

$140 online, $170 nearby

Reviews

Summary - Based on 159 reviews

What people are saying

pictures  "We use the product to take quickly photos."
features  "Impressive panoramic feature."
zoom/lens "It also record better and focus better on sunny days."
design  "It has the slightest grip but it’s sufficient."
video  "Video zoom is choppy."
battery life "Even better, the battery lasts long."
screen  "I Love the Sony’s 3” screen which I really wanted."
Not just one problem

- \((e_j, f_{jk}, s_{ijkl}, h_i, t_l)\),
  - \(e_j\) - a target entity: Named Entity Extraction (more)
  - \(f_{jk}\) - a feature/aspect of \(e_j\): Information Extraction
  - \(s_{ijkl}\) is sentiment: Sentiment Identification
  - \(h_i\) is an opinion holder: Information/Data Extraction
  - \(t_l\) is the time: Information/Data Extraction
  - 5 pieces of information must match

Natural language processing issues
  - Coreference resolution
  - Synonym match (voice = sound quality)
  - …
Highly researched sub-problems

- **Document-level**
  - Classify reviews as positive or negative

- **Sentence-level**
  - Subjectivity and sentiment classification, **but note**
    - both subjective & objective sentences can have opinion.
    - Many subjective sentences have no +ve or –ve opinion

- **Aspect-level sentiment analysis**
  - Aspect extraction
  - Aspect sentiment classification

- **A key challenge is about discovery**
Entity discovery/extraction

- Given BMW and Ford, find all car brands and models and different ways of writing them in a text collection
  - Although similar, it is different from the traditional named entity recognition (NER).

**Formulation**: Given a set \( Q \) of seed entities of a particular class \( C \), and a set \( D \) of candidate entities, we wish to determine which of the entities in \( D \) belong to \( C \).

- A classification problem. It needs a binary decision for each entity in \( D \) (belonging to \( C \) or not)
  - But it’s normally solved as a ranking problem
Some methods  (Li et al 2010, Zhang and Liu, 2011)

- **Distributional similarity**: This is the traditional method used in NLP, which compare the surrounding text of candidates.
  - It performs poorly.
- **PU learning**: learning from positive and unlabeled examples.
  - S-EM algorithm (Liu et al. 2002)
- **Bayesian Sets**: We extended the method given in (Ghahramani and Heller, NIPS-05).
Determine sentiment is hard!

- Most algorithms use sentiment terms and/or classification to determine sentiments.
  - Sentiment terms do not go very far.
- There is a long tail of cases that sentiment terms cannot handle
  - There seem to be a unlimited number of ways that one can use to express opinions
    - Every domain has some peculiar cases, which make the general opinion mining very hard in practice.
- We need a lot of knowledge discovery
Some Example Sentences

- I am so happy because my new iPhone is nothing like my old ugly Nokia phone.
- After my wife and I slept on the mattress for a week, I found a hill in the middle.
- Since I had a lot of pain on my back, so my doctor put me on the drug, and only two days after, I have no more pain.
- After taking the drug, my blood pressure went to 400.
- Trying out Google chrome because Firefox keeps crashing.
- Anyone know a good Sony camera?
- Anyone know how to fix this lousy washer?
- If I can find a good Sony camera, I will buy it.
- If you are in for a good camera, go for Canon S500.
- What a great car, it stopped working in the second day.
Opinions/sentiments are governed by many rules, e.g.,

- **Opinion word or phrase**, ex: “This is a good car”
  - $P ::= \text{a positive opinion word or phrase}$
  - $N ::= \text{an negative opinion word or phrase}$

- **Desirable or undesirable facts**, ex: “After my wife and I slept on it for two weeks, I noticed a mountain in the middle of the mattress”
  - $P ::= \text{desirable fact}$
  - $N ::= \text{undesirable fact}$
Basic rules of opinions

- High, low, increased and decreased quantity of a positive or negative potential item, ex: “The battery life is long.”

- \( PO \) ::= no, low, less or decreased quantity of NPI
  | large, larger, or increased quantity of PPI

- \( NE \) ::= no, low, less, or decreased quantity of PPI
  | large, larger, or increased quantity of NPI

- \( NPI \) ::= a negative potential item

- \( PPI \) ::= a positive potential item
Basic rules of opinions

- **Decreased and increased quantity of an opinionated item, ex:** “This drug reduced my pain significantly.”
  
  \[ \text{PO} ::= \text{less or decreased N} \mid \text{more or increased P} \]
  
  \[ \text{NE} ::= \text{less or decreased P} \mid \text{more or increased N} \]

- **Deviation from the desired value range:** “This drug increased my blood pressure to 200.”
  
  \[ \text{PO} ::= \text{within the desired value range} \]
  
  \[ \text{NE} ::= \text{above or below the desired value range} \]
Basic rules of opinions

- Producing and consuming resources and wastes, ex: “This washer uses a lot of water”

PO ::= produce a large quantity of or more resource
| produce no, little or less waste
| consume no, little or less resource
| consume a large quantity of or more waste

NE ::= produce no, little or less resource
| produce some or more waste
| consume a large quantity of or more resource
| consume no, little or less waste
Desirable or undesirable facts
(Zhang and Liu, 2011)

“After sleeping on the mattress for one month, a valley has formed in the middle.”

In most sentiment analysis task, we need opinion words, e.g., good, bad, hate, crap, junk, etc.

But objective nouns indicating desirable and undesirable facts can imply opinions too.

E.g., How to discover such nouns from a domain corpus?
The technique

- Sentiment analysis to determine whether the context is +ve or –ve.
  - E.g., “I saw a valley in two days, which is terrible.”
  - This is a negative context.
- Statistical test to find +ve and –ve candidates.

\[ Z = \frac{p - p_0}{\sqrt{p_0(1-p_0) / n}} \]

- Pruning to move those unlikely ones through sentiment homogeneity.
Pruning

- For an aspect with an implied opinion, it has a fixed opinion, either +ve or –ve, but not both.
- We find two direct modification relations using a dependency parser.
  - Type 1: \( O \rightarrow O\text{-Dep} \rightarrow A \)
    - e.g. “This TV has good picture quality.”
  - Type 2: \( O \rightarrow O\text{-Dep} \rightarrow H \leftarrow A\text{-Dep} \leftarrow A \)
    - e.g. “The springs of the mattress are bad.”
- If an aspect has mixed opinions based on the two dependency relations, prune it.
Opinions implied by resource usage
(Zhang and Liu, 2011)

- Resource usage descriptions often imply opinions (as mentioned in rules of opinions)
  - E.g., “This washer uses a lot of water.”

- Two key roles played by resources usage:
  - An important aspect of an entity, e.g., water usage.
  - Imply a positive or negative opinion

- Resource usages that imply opinions can often be described by a triple.
  - (verb, quantifier, noun_term),
  - Verb: uses, quantifier: “a lot of “, noun_term: water
The proposed technique

- The proposed method is graph-based.
  - Stage 1: Identifying Some Global Resource Verbs
    - Identify and score common resource usage verbs used in almost any domain, e.g., “use” and “consume”
  - Stage 2: Discovering Resource Terms in each Domain Corpus
    - Use a graph-based method considering occurrence probabilities.
    - With resource verbs identified from stage 1 as the seeds.
    - Score domain specific resource usage verbs and resource terms.
The algorithm

**Algorithm:** MRE \((Q, G)\)

**Input:** A global resource verb set \(Q\) with their hub scores computed from HITS in stage 1, and \(G\) is the bipartite graph

**Output:** a ranked list of candidate resource terms

1. \(u^0(i) \leftarrow H(i)\) of verb \(i\), if verb \(i \in Q\)
2. \(u^0(i) \leftarrow \arg\min_{r \in Q} \{H(r)\}\), if verb \(i \notin Q\)
3. **Repeat** till convergence
4. \(r^{n+1}(j) = \sum_{(i,j) \in L} p_{ij} u^n(i)\)
5. \(u^{n+1}(i) = \sum_{(i,j) \in L} p_{ji} r^n(j)\)
6. normalize \(r(j)\) and \(u(i)\)
7. Output the ranked candidate resource terms based on their \(r(j)\) score values.
Coreference resolution: semantic level?

- Coreference resolution (Ding and Liu, 2010)
  - “I bought the Sharp tv a month ago. The picture quality is so bad. Our other Sony tv is much better than this Sharp. *It is also so expensive*.”
    - “it” means “Sharp”
  - “I bought the Sharp tv a month ago. The picture quality is so bad. Our other Sony tv is much better than this Sharp. *It is also very reliable*.”
    - “it” means “Sony”

- Sentiment consistency.
Coreference resolution (contd)

- “The picture quality of this Canon camera is very good. *It* is not expensive either.”
  - Does “it” mean “Canon camera” or “Picture Quality”?
    - Clearly it is Canon camera because picture quality cannot be expensive.
    - Commonsense knowledge, but can be discovered.

- For coreference resolution, we actually need to
  - do sentiment analysis first, and
  - mine adjective-noun associations using dependency

- Finally, use supervised learning
Comparative Opinions
(Jindal and Liu, 2006)

- **Gradable**
  - **Non-Equal Gradable**: Relations of the type *greater* or *less than*
    - Ex: “optics of camera A is better than that of camera B”
  - **Equative**: Relations of the type *equal to*
    - Ex: “camera A and camera B both come in 7MP”
  - **Superlative**: Relations of the type *greater* or *less than all others*
    - Ex: “camera A is the cheapest in market”
**Analyzing Comparative Opinions**

- **Objective**: Given an opinionated document $d$, extract comparative opinions:

  $$(E_1, E_2, F, po, h, t),$$

  where $E_1$ and $E_2$ are the entity sets being compared based on their shared features/aspects $F$, $po$ is the preferred object set of the opinion holder $h$, and $t$ is the time when the comparative opinion is expressed.

- **Note**: not positive or negative opinions.
Deal with comparative opinions

- Gradable comparative sentences can be dealt with *almost* as normal opinion sentences.
  - E.g., “optics of camera A is better than that of camera B”
  - Positive: “optics of camera A”
  - Negative: “optics of camera B”

- **Difficulty**: recognize non-standard comparatives
  - E.g., “I am so happy because my new iPhone is nothing like my old slow ugly Droid.”
  - ?
Some techniques (Jindal and Liu, 2006, Ding et al, 2009)

- Identify comparative sentences
  - Using class sequential rules as attributes in the data, and then
  - Supervised learning

- Extraction of different items
  - Label sequential rules
  - Conditional random fields

- Determine opinion orientations
  - Parsing and opinion lexicon
    - We have not used supervised learning
Group aspects synonyms (Zhai et al. 2011a, b)

- Once aspects expressions are discovered, group them into /aspect categories.
  - Power usage and battery life are the same.

- A variety of information is used in clustering
  - Lexical similarity based on WordNet
  - Distributional information
  - Syntactical information/constraints

- Two Methods:
  - Clustering: EM-based method.
The EM-based method

- **WordNet similarity**

\[
Jcn(w_1, w_2) = \frac{1}{IC(w_1) + IC(w_2) - 2 \times Res(w_1, w_2)}
\]

- **EM-based probabilistic clustering**

\[
P(w_t | c_j) = \frac{1 + \sum_{i=1}^{|D|} N_{ti} P(c_j | d_i)}{|V| + \sum_{m=1}^{|V|} \sum_{i=1}^{|D|} N_{mi} P(c_j | d_i)}
\]

\[
P(c_j) = \frac{1 + \sum_{i=1}^{|D|} P(c_j | d_i)}{|C| + |D|}
\]

\[
P(c_j | d_i) = \frac{P(c_j) \prod_{k=1}^{|d_i|} P(w_{d_i,k} | c_j)}{\sum_{r=1}^{|C|} P(c_r) \prod_{k=1}^{|d_i|} P(w_{d_i,k} | c_r)}
\]
Constrained Topic Modeling

- **Constrained topic model**: Constrained-LDA
- In topic modeling, we add probabilistic constraints
  - Must-links
  - Cannot link
- In Gibbs sampling, we consider constraints to guide its topic assignments of aspect terms.
Find evaluative opinions in discussions
(Zhai et al. 2011)

- Existing research focuses on product reviews
  - reviews are opinion-rich and
  - contain little irrelevant information.

- Not true about online discussions.
  - Many of the postings do not express opinions about the discussion topic.
    - **Evaluative opinions**, “The German defense is strong.”
    - **Non-evaluative opinions**, “I feel so sad for Argentina.” “you know nothing about defense”

- **Goal**: discover evaluative opinion sentences.
3. The Proposed Technique

- **Intuitions:** (1) An *evaluative* opinion should comment on a topic/entity or some aspects of it. (2) *Evaluation words* and *emotion words* are indications of evaluative and emotional sentences, respectively.

- **Overview:** Given the raw discussion postings, the algorithm works in 4 steps to identify *evaluative* sentences.
3.1 Extraction of Aspects and Expansion of Evaluation and Emotion Lexicons

Input: Text corpus \( R \); Evaluation word seeds \( vas \); Emotion word seeds \( mos \). // Not sufficient

Output: All evaluation words \( VA \); All emotion words \( MO \); All aspects: \( A \)

Task 1. Extract aspects using evaluation/emotion words;
Task 2. Extract aspects using extracted aspects;
Task 3. Extract evaluation words and emotion words using the given or extracted evaluation words and emotion words respectively.
Double-Propagation (DP)

- We use the Double Propagation method in (Qiu et al 2009; 2011).
- The idea is that an opinion has a target.
  - Ex: This Sony camera is great.
- This technique needs a dependency parser.
- In this work, we are interested in Chinese microblog (weibo) discussions
  - But Chinese dependency parsers are not accurate.
- We approximate the DP method using POS tags
3.2 Aspects, Evaluation Words and Emotion Words Interaction

- An extracted aspect that is associated with many *evaluation words* is more likely to indicate an evaluative sentence. Then, we want to give a high score to the aspect.

- An extracted aspect that is associated with many *emotion words* is not a good indicator of an evaluative sentence. It should be assigned a low score.
3.2 Aspects, Evaluation Words and Emotion Words Interaction

- An evaluation word that does not modify good (high scored) aspects are likely to be a wrong evaluation word, and should be weighted down.

\[ \text{asp}(a_i) = \lambda \sum_{(i,j) \in E_{va-a}} \text{eva}(va_j) - (1 - \lambda) \sum_{(i,k) \in E_{mo-a}} \text{emo}(mo_k) \]  

\[ \text{eva}(va_j) = \sum_{(i,j) \in E_{va-a}} \text{asp}(a_i) \]  

- The more evaluative the aspects are, the less emotional their associated emotion words should be.

\[ \text{tmp}(mo_k) = \sum_{(i,k) \in E_{mo-a}} \text{asp}(a_i) \]

\[ \text{emo}(mo_k) \propto -\text{tmp}(mo_k) \]

\[ \text{emo}(mo_k) = -\text{tmp}(mo_k) + \max = \max - \text{tmp}(mo_k) \]

\[ \max = \max\{\text{tmp}(mo_1), \text{tmp}(mo_2), ..., \text{tmp}(mo_{|v_{mo}|})\} \]
Opinion mining or sentiment analysis is a fascinating NLP or text mining problem.

It is also restricted NLP problem
- Because we only need to understand one aspect of the semantic meaning.

General NLP is probably hopeless.

But can we solve this restricted problem?
- Although many challenges, there are already numerous applications.
- I am optimistic.
See my page and the book:

- [http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html](http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html)