

CSCI5070 Advanced Topics in Social Computing

Sentiment Analysis

Irwin King

The Chinese University of Hong Kong

king@cse.cuhk.edu.hk

©2012 All Rights Reserved.



Introduction

- Two main types of textual information
 - Facts and Opinions
- Most current text information processing methods (e.g., web search, text mining) work with factual information.
- Sentiment analysis or opinion mining
 - Computational study of opinions, sentiments and emotions expressed in text



Introduction

- Importance of opinions
 - Opinions are important because whenever we need to make a decision, we want to hear others' opinions
 - In the past,
 - **Individuals**: opinions from friends and family
 - **Businesses**: surveys, focus groups, consultants ...
- Word-of-mouth on the Web
 - **User-generated media**: One can express opinions on anything in reviews, forums, discussion groups, blogs ...
 - **Opinions of global scale**: No longer limited to:
 - **Individuals**: one' s circle of friends
 - **Businesses**: Small scale surveys, tiny focus groups, etc.



A Fascinating Problem!

- Intellectually challenging & major applications
 - A popular research topic in recent years in NLP and Web data mining
 - 20–60 companies in USA alone
- It touches every aspect of NLP and yet is restricted and confined
 - Little research in NLP/Linguistics in the past
- Potentially a major technology from NLP
 - But “not yet” and not easy
 - Data sourcing and data integration are hard too



An Example Review

- Two main types of textual information

“I bought an iPhone a few days ago. It was such a nice phone. The touch screen was really cool. The voice quality was clear too. Although the battery life was not long, that is ok for me. 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, and wanted me to return it to the shop. ...”

- **What do we see?**

- Opinions, targets of opinions, and opinion holders



What is an Opinion?

- An opinion is a **quintuple**

$$(o_j, f_{jk}, so_{ijkl}, h_i, t_l),$$

where

- o_j is a target object
- f_{jk} is a feature of the object o_j .
- so_{ijkl} is the sentiment value of the opinion of the opinion holder h_i on feature f_{jk} of object o_j at time t_l . so_{ijkl} is +ve, -ve, or neu, or a more granular rating.
- h_i is an opinion holder.
- t_l is the time when the opinion is expressed.



Sentiment Analysis Tasks

- Task 1. Sentiment identification (Subjectivity identification): identify whether a piece of text expresses opinions
- Task 2. Sentiment orientation classification: determine the orientation of an opinionated text



Sentiment Analysis Levels

- *Document-level*: identify if the **document** (e.g. product reviews, blogs, forum posts) expresses opinions and whether the opinions are positive, negative, or neutral
- *Sentence-level*: identify if a **sentence** is opinionated and whether the opinion is positive, negative, or neutral
- *Attribute-level*: extract the **object attributes** (e.g. image quality, zoom size) that are the subject of an opinion and the opinion orientations
- As the object becomes more granular, the intensity /difficulty increases



Document-level Sentiment

Analysis

- Tasks: identify if the document expresses opinions and if yes classify the document into positive, negative, or neutral based on the overall sentiments expressed by opinion holders
- Assumptions
 - the document is opinionated on a single object
 - the opinions are from a single opinion holder
- Similar to but different from topic-based text classification
 - In topic-based text classification, topic words are important
 - In sentiment classification, **opinion words** are more important, e.g. wonderful, fabulous, terrible



A Product Review Example

“My Canon Powershot takes **great** pictures! ... My friend had gotten one about a year ago and **she loves it**. So, after seeing her enthusiasm about it I decided to get one and **I will never go back to any other camera**. **I absolutely love** this camera. I believe that **every person on Earth should own one of these**. ... It is **amazing!** ... There is **not** one thing I **hate** about this product, which is strange because I am a very picky person! ...”



Opinion Words

- Also known as polarity words, sentiment words, opinion lexicon, or opinion-bearing words, e.g.
- Positive: wonderful, elegant, amazing
- Negative: horrible, disgusting, poor
- Base type (examples above) and comparative type (e.g. better, worse)
- How to generate them
 - **Manually crafted**
 - **Automatic learning from patterns**



A Simple Method - Counting Opinion Words

- Opinion/polarity words: dominating indicators of sentiments, especially adjectives, adverbs, and verbs, e.g. “I absolutely love this camera. It is amazing!” .
- Pre-defined opinion words: good, terrible, ...
- Assign orientation score (+1, -1) to all words
 - Positive opinion words (+1): great, amazing, love
 - Negative opinion words (-1): horrible, hate
 - Strength value [0, 1] can be used too
- The orientation score of the document is the sum of orientation scores of all opinion words found

- The Chinese University of Hong Kong, GSC15070 Advanced Topics in Social Computing, Irwin King 1 = 3



Rule-based Method

- Is simply counting opinion words good enough? No!
 - “There is **not** one thing I **hate** about this product”
- We need to handle negation: “not ... hate” implies like
 - Simple rules can be manually created
 - “not ... negative” **positive**
 - “never ... negative” **positive**
 - The previous review has a score of $4 + 1 = 5$
 - Note: negation needs to be handled with care, e.g. “not” in “not only ... but also” does not change the orientation



Limitation of Rule-based System

- An expensive task
 - Only a limited number of opinion words can be found
 - Only a limited number of patterns can be created



Other Approches

- Dictionary-based Approach - Bootstrapping
- Corpus-based Approach
- Supervised Learning
 - Feature Extraction
 - Classification
 - Naïve Bayes
 - Maximum Entropy
 - Support Vector Machines
 - Logistic Regression



Declaration

- The slides are borrowed from
 - <http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>
 - http://web.cs.dal.ca/~yongzhen/publication/paper/ICDM2011_SentimentAnalysisInPracticeTutorial.pdf

