AI6122 Text Data Management & Analysis

Topic: Sentiment Analysis

A high level overview

- What is sentiment analysis

- Applications

- Tasks
  - Document sentiment classification
  - Sentence subjectivity & sentiment classification
  - Aspect-based sentiment analysis
  - Sentiment lexicon

Sentiment Analysis vs Opinion Mining

- Computational study of opinions, sentiments, evaluations, attitudes, appraisal, affects, views, emotions, subjectivity, etc., expressed in text.
  - Reviews, blogs, discussions, news, comments, feedback, or any other documents.

- Terminology: Sentiment Analysis vs Opinion Mining
  - Basically means the same
  - Sentiment analysis is more widely used in industry.
  - Both are widely used in academia

Why sentiment analysis?

- Movie
  - is this review positive or negative?

- Products
  - what do people think about the new iPhone?

- Public sentiment
  - how is consumer confidence? Is despair increasing?

- Politics
  - what do people think about this candidate or issue?

- Prediction
  - predict election outcomes or market trends from sentiment

Applications

- Businesses and organizations
  - Benchmark products and services; market intelligence.
  - Businesses spend a huge amount of money to find consumer opinions using consultants, surveys and focus groups, etc

- Individuals
  - Make decisions to buy products or to use services
  - Find public opinions about political candidates and issues

- Ads placements: Place ads in the social media content
  - Place an ad if one praises a product.
  - Place an ad from a competitor if one criticizes a product.

- Opinion retrieval: provide general search for opinions.

Scherer Typology of Affective States

- Emotion: brief organically synchronized … evaluation of a major event
  - angry, sad, joyful, fearful, ashamed, proud, elated

- Mood: diffuse non-caused low-intensity long-duration change in subjective feeling
  - cheerful, gloomy, irritable, listless, depressed, buoyant

- Interpersonal stances: affective stance toward another person in a specific interaction
  - friendly, flirtatious, distant, cold, warm, supportive, contemptuous

- **Attitudes: enduring, affectively colored beliefs, dispositions towards objects or persons**
  - **liking, loving, hating, valuing, desiring**

- Personality traits: stable personality dispositions and typical behavior tendencies
  - nervous, anxious, reckless, morose, hostile, jealous

Sentiment Analysis

- Sentiment analysis is the detection of **attitudes**

Holder (source)	→ Attitude →	Target (source)

- Type of attitude
  - From a set of types: Like, love, hate, value, desire, etc.
  - More commonly weighted polarity: positive, negative, neutral, together with strength

- Text containing the attitude
  - Sentence or entire document

NANYANG TECHNOLOGICAL UNIVERSITY | SINGAPORE

Sentiment Analysis

- Simplest task:
  – Is the attitude of this text positive or negative?

- More complex:
  – Rank the attitude of this text from 1 to 5

- Advanced:
  – Detect the target, source, or complex attitude types

Many tasks in more detailed view

- Document sentiment classification
- Sentence subjectivity & sentiment classification

- Aspect-based sentiment analysis
- Aspect-based opinion summarization

- Sentiment lexicon generation
- Mining comparative opinions

- Opinion spam detection
- Utility or helpfulness of reviews

What do we look for in Sentiment Analysis

- Example review:
  - "*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, …*"

- Sentiment analysis
  - Document level, i.e., is this review + or -?
  - Sentence level, i.e., is each sentence + or -?
  - Entity and feature/aspect level

  - Opinion targets: entities and their features/aspects
  - Sentiments: positive and negative
  - Opinion holders: persons who hold the opinions
  - Time: when opinions are expressed

Problem Statement

- Opinion definition. What is an opinion?
  - Can we provide a structured definition?
  - If we cannot structure a problem, we probably do not understand the problem.

- Opinion summarization
  - Opinions are subjective. An opinion from a single person is often not sufficient for action.
  - We need opinions from many people, and thus opinion summarization.

Two main types of opinions

- Regular opinions:
  - Sentiment/opinion expressions on some target entities
  - Direct opinions: "The touch screen is really cool."
  - Indirect opinions: "After taking the drug, my pain has gone."

- A regular opinion has the follow basic components
  - a **target**: to which the opinion is made on
  - **sentiment value** of the opinion: e.g., positive, negative or neutral
  - **opinion holder**: who makes the opinion
  - a time: when the opinion was made

- Comparative opinions: Comparisons of more than one entity.
  - E.g., "iPhone is better than Blackberry."

Opinion 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, …*"

Feature Based Summary of 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

Opinion mining in the real world

- Source the data: e.g., reviews, blogs, tweets..
  - Crawl all data, store and search them, or
  - Crawl only the target data

- Extract the right entities & aspects
  - Group entity and aspect expressions,
  - e.g., Moto = Motorola, photo = picture

- Aspect-based opinion mining (sentiment analysis)
  - Discover all necessary quantities e.g., opinion holder, target, sentiment value…
  - Store them in some database or indexes

- Aspect based opinion summary

Many tasks in more detailed view

- **Document sentiment classification**
- Sentence subjectivity & sentiment classification

- Aspect-based sentiment analysis
- Aspect-based opinion summarization

- Sentiment lexicon generation
- Mining comparative opinions

- Opinion spam detection
- Utility or helpfulness of reviews

Sentiment classification

- Classify a whole opinion document (e.g., a review) based on the overall sentiment of the opinion holder
  - Classes: Positive, negative (possibly neutral)
  - Neutral or no opinion is hard (many papers ignore it).

- It is basically a **text classification** problem
  - In topic-based text classification (e.g., computer, sport, science), topic words are important.
  - In sentiment classification, opinion words are more important, e.g., great, excellent, horrible, bad, worst, etc.

- **Opinion/sentiment words**
  - Words and phrases that express desired or undesired states or qualities.

Unsupervised classification (Turney, 2002)

- Data: reviews from epinions.com on automobiles, banks, movies, and travel destinations.

- Step 1: Part-of-speech (POS) tagging
  - Extracting two consecutive words (two-word phrases) from reviews if their tags conform to some given patterns, e.g., (1) JJ, (2) NN.

	First word	Second word	Third word (Not Extracted)
1.	JJ	NN or NNS	anything
2.	RB, RBR, or RBS	JJ	not NN nor NNS
3.	JJ	JJ	not NN nor NNS
4.	NN or NNS	JJ	not NN nor NNS
5.	RB, RBR, or RBS	VB, VBD, VBN, or VBG	anything

Unsupervised classification (Turney, 2002)

- Step 2: Estimate the sentiment orientation (SO) of the extracted phrases
  - Use Pointwise mutual information

$$PMI(w_1, w_2) = \log(\frac{P(w_1, w_2)}{P(w_1)P(w_2)})$$

  - Semantic orientation (SO)
    - SO(phrase) = PMI(phrase, "excellent") - PMI(phrase, "poor")

- Step 3: Compute the average SO of all phrases
  - classify the review as positive if average SO is positive, negative otherwise.

Supervised learning

- Directly apply supervised learning techniques to classify reviews into positive and negative.
  - Like a text classification problem

- Many classification techniques have been evaluated
  - Naïve Bayes
  - Maximum entropy
  - Support vector machines

- Features (document representation)
  - negation tag, unigram (single words), bigram, POS tag, position.

- Categories (for Movie reviews with star ratings)
  - 4-5 stars as positive
  - 1-2 stars as negative

Supervised learning for document level SA

- Probably the most extensive studied problem
  - Including domain adaption and cross-lingual, etc.

- Key: feature engineering. A large set of features have been tried by researchers.
  - Terms frequency and different IR weighting schemes
  - Part of speech (POS) tags
  - Opinion words and phrases
  - Negations
  - Syntactic dependency

Sentence-level sentiment classification

- Document-level sentiment classification is too coarse for most applications.

- Sentence level analysis usually consist of two steps
  - Subjectivity classification: subjective vs objective sentences
  - Sentiment classification of subjective sentences: positive vs negative
  - Most techniques use supervised learning.

- Assumption:
  - Each sentence expresses a single positive or negative opinion/sentiment.
  - Good for simple sentence like: "I like this car";
  - Not true for compound and "complex" sentences
    - "I like the picture quality but battery life sucks."
    - "Apple is doing very well in this lousy economy."

Aspect-level Sentiment Analysis

- Sentiment classification at both the document and sentence (or clause) levels are useful, but
  - They do not find what people liked and disliked.
  - They do not identify the targets of opinions, i.e., entities and their aspects

- Aspect-based opinion mining and summarization
  - Much of the research is based on online reviews
  - For reviews, aspect-based sentiment analysis is easier because the entity (i.e., product name) is usually known
  - Reviewers simply express positive and negative opinions on different aspects of the entity.
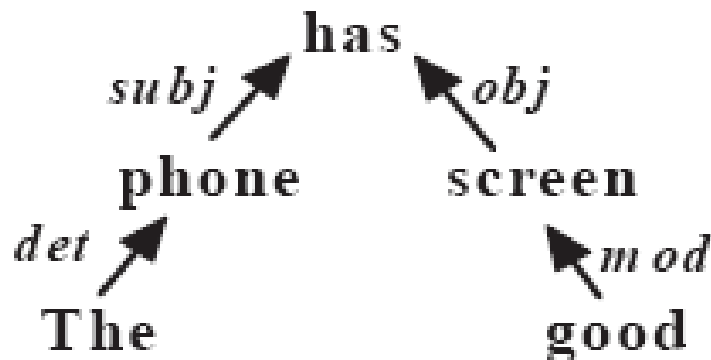
Aspect extraction

- Goal: Given an opinion corpus, extract all aspects

- A frequency-based approach
  - nouns (NN) that are frequently talked about are likely to be true aspects (called frequent aspects) .

- Why the frequency-based approach?
  - Different reviewers tell different stories (irrelevant)
  - When product aspects/features are discussed, the words they use converge.
  - They are the main aspects.

- Sequential/association pattern mining finds frequent nouns and noun phrases.

Aspect extraction by Dependency

- An opinion should have a target, entity or aspect.
  - Use dependency of opinions & aspects to extract both aspects & opinion words.
  - Knowing one helps find the other, e.g., "*The rooms are spacious*"

- Dependency helps to extract both aspects and opinion words.
  - A domain independent method.

Sentiment Lexicon

- The General Inquirer
  - Home page: http://www.wjh.harvard.edu/~inquirer
  - List of Categories: http://www.wjh.harvard.edu/~inquirer/homecat.htm
  - Spreadsheet: http://www.wjh.harvard.edu/~inquirer/inquirerbasic.xls

- Categories:
  - Positive (1915 words) and Negative (2291 words)
  - Strong vs Weak, Active vs Passive, Overstated versus Understated
  - Pleasure, Pain, Virtue, Vice, Motivation, Cognitive Orientation, etc

- Free for Research Use

Philip J. Stone, Dexter C Dunphy, Marshall S. Smith, Daniel M. Ogilvie. 1966. The General Inquirer: A Computer Approach to Content Analysis. MIT Press

NANYANG TECHNOLOGICAL UNIVERSITY | **SINGAPORE**

Sentiment Lexicon

- LIWC (Linguistic Inquiry and Word Count)
  - Home page: http://www.liwc.net/
  - 2300 words, >70 classes

- Affective Processes
  - negative emotion (bad, weird, hate, problem, tough)
  - positive emotion (love, nice, sweet)

- Cognitive Processes
  - Tentative (maybe, perhaps, guess), Inhibition (block, constraint)

- Pronouns, Negation (no, never), Quantifiers (few, many)
- $30 or $90 fee

Pennebaker, J.W., Booth, R.J., & Francis, M.E. (2007). Linguistic Inquiry and Word Count: LIWC 2007. Austin, TX

Sentiment Lexicon

- MPQA Subjectivity Cues Lexicon
  - Home page: http://www.cs.pitt.edu/mpqa/subj_lexicon.html

- 6885 words from 8221 lemmas
  - 2718 positive
  - 4912 negative

- Each word annotated for intensity (strong, weak)

- GNU GPL

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. Proc. of HLT-EMNLP-2005.

Riloff and Wiebe (2003). Learning extraction patterns for subjective expressions. EMNLP-2003.

NANYANG TECHNOLOGICAL UNIVERSITY | SINGAPORE

Sentiment Lexicon

- Bing Liu Opinion Lexicon
  - [Bing Liu's Page on Opinion Mining](#)
  - http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar
  - 6786 words: 2006 positive, 4783 negative

- SentiWordNet
  - Home page: http://sentiwordnet.isti.cnr.it/
  - All WordNet synsets automatically annotated for degrees of positivity, negativity, and neutrality/objectiveness

Minqing Hu and Bing Liu. Mining and Summarizing Customer Reviews. ACM SIGKDD-2004.

NANYANG TECHNOLOGICAL UNIVERSITY | SINGAPORE

Many tasks in Sentiment Analysis

- **Document sentiment classification**
- **Sentence subjectivity & sentiment classification**

- **Aspect-based sentiment analysis**
- Aspect-based opinion summarization

- **Sentiment lexicon** generation
- Mining comparative opinions

- Opinion spam detection
- Utility or helpfulness of reviews