Sentiment Analysis and Opinion Mining

Introduction

- **Opinion mining or sentiment analysis**
  - Computational study of opinions, sentiments, subjectivity, evaluations, attitudes, appraisal, affects, views, emotions, etc., expressed in text.
    - Reviews, blogs, discussions, news, comments, feedback, or any other documents
- Terminology:
  - Sentiment analysis is more widely used in industry.
  - Both are widely used in academia
- But they can be used interchangeably.

Why are opinions 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. In the past,

  - Individuals: seek opinions from friends and family

  - Organizations: use surveys, focus groups, opinion polls, consultants.

Introduction – social media + beyond

- **Word-of-mouth on the Web**
  - Personal experiences and opinions about anything in reviews, forums, blogs, Twitter, micro-blogs, etc
  - Comments about articles, issues, topics, reviews, etc.
  - Postings at social networking sites, e.g., facebook.
- **Global scale:** No longer – one's circle of friends
- **Organization internal data**
  - Customer feedback from emails, call centers, etc.
- **News and reports**
  - Opinions in news articles and commentaries

Introduction – 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.

A fascinating problem!

- **Intellectually challenging & many applications.**
  - A popular research topic in NLP, text mining, and Web mining in recent years (Shanahan, Qu, and Wiebe, 2006 (edited book); Surveys - Pang and Lee 2008; Liu, 2006 and 2011; 2010)
  - It has spread from computer science to management science (Hu, Pavlou, Zhang, 2006; Archak, Ghose, Ipeirotis, 2007; Liu Y, et al 2007; Park, Lee, Han, 2007; Dellarocas, Zhang, Awad, 2007; Chen & Xie 2007).
  - 40-60 companies in USA alone
- It touches every aspect of NLP and yet is confined.
  - Little research in NLP/Linguistics in the past.
- Potentially a major technology from NLP.
  - But it is hard.

A large research area

- **Many names and tasks** with somewhat different objectives and models

  - Sentiment analysis
  - Opinion mining
  - Sentiment mining
  - Subjectivity analysis
  - Affect analysis
  - Emotion detection
  - Opinion spam detection
  - *Etc.*

Roadmap

- **Opinion Mining Problem**
- Document sentiment classification
- Sentence subjectivity & sentiment classification
- Aspect-based sentiment analysis
- Aspect-based opinion summarization
- Opinion lexicon generation
- Mining comparative opinions
- Some other problems
- Opinion spam detection
- Utility or helpfulness of reviews
- Summary

Structure the unstructured (Hu and Liu 2004)

- **Structure the unstructured**: Natural language text is often regarded as unstructured data.
- The problem definition should provide a structure to the unstructured problem.
  - Key tasks: Identify key tasks and their inter-relationships.
  - Common framework: Provide a common framework to unify different research directions.
  - Understanding: help us understand the problem better.

Problem statement

- **It consists of two aspects of abstraction**

(1) 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.

(2) Opinion summarization. why?
  - Opinions are subjective. An opinion from a single person (unless a VIP) is often not sufficient for action.
  - We need opinions from many people, and thus opinion summarization.