SENTIMENT ANALYSIS FUNDAMENTALS

INTRODUCTION

DR LEK HSIANG HUI

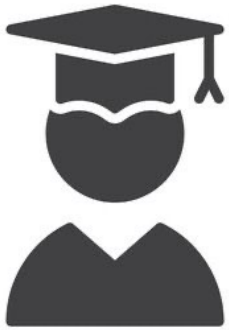
About Me

- LEK Hsiang Hui
- Senior Lecturer
- Department of Information Systems and Analytics

- Teaches Courses on:
  - Analytics
  - Software Engineering

- Undergraduates, Graduates, and Executives

About Me



Did PhD in the area Natural Language Processing (Sentiment Analysis)

about us

Based in Singapore's silicon valley, Blk 73 Launchpad, we are a big data analytics company, revolutionizing global and local brand analytics ranking and profiling. With billions of online conversations, we have the data and technologies to determine any brands' relative brand position globally. We have the ability to automatically conduct extensive user profiling for brands wanting to know more about their customers' preferences and desires. Through workshops and trainings, we help brands make sense of data and get them started with digital marketing.

what we offer

CUSTOMER INSIGHTS

Want to know what netizens are talking about your brand, competitors or a brand? How can you improve your services or business offerings?

INDUSTRY KNOWLEDGE

What is the latest trend happening - for your brand, your industry or target audience? What can we learn from others in your industry to help you in your marketing strategy?

INFLUENCER PROFILING

Who are your influencers? Can you turn them into leads? Or use them to bring in leads?

Data made simple

STANDICES

how we work + how we can help you

REPORTS

With our technology and the immense data that are available on the web, we can provide customer insights, industry knowledge and influencer profiling.

TRAININGS

We conduct trainings and workshops to help brands solve problems using their data and to provide knowledge to propel your marketing strategies. Let us help you understand analytics and marketing tactics.

TALK TO US

to let us do the thinking!

email: enquiry@standices.com

Went on to co-founded a Big Data Analytics Company

Contact

https://www.facebook.com/hsianghui

@hsianghui

https://www.linkedin.com/in/hsianghui/

hsianghui@nus.edu.sg

Introduction to Text Mining

What is Text Mining?

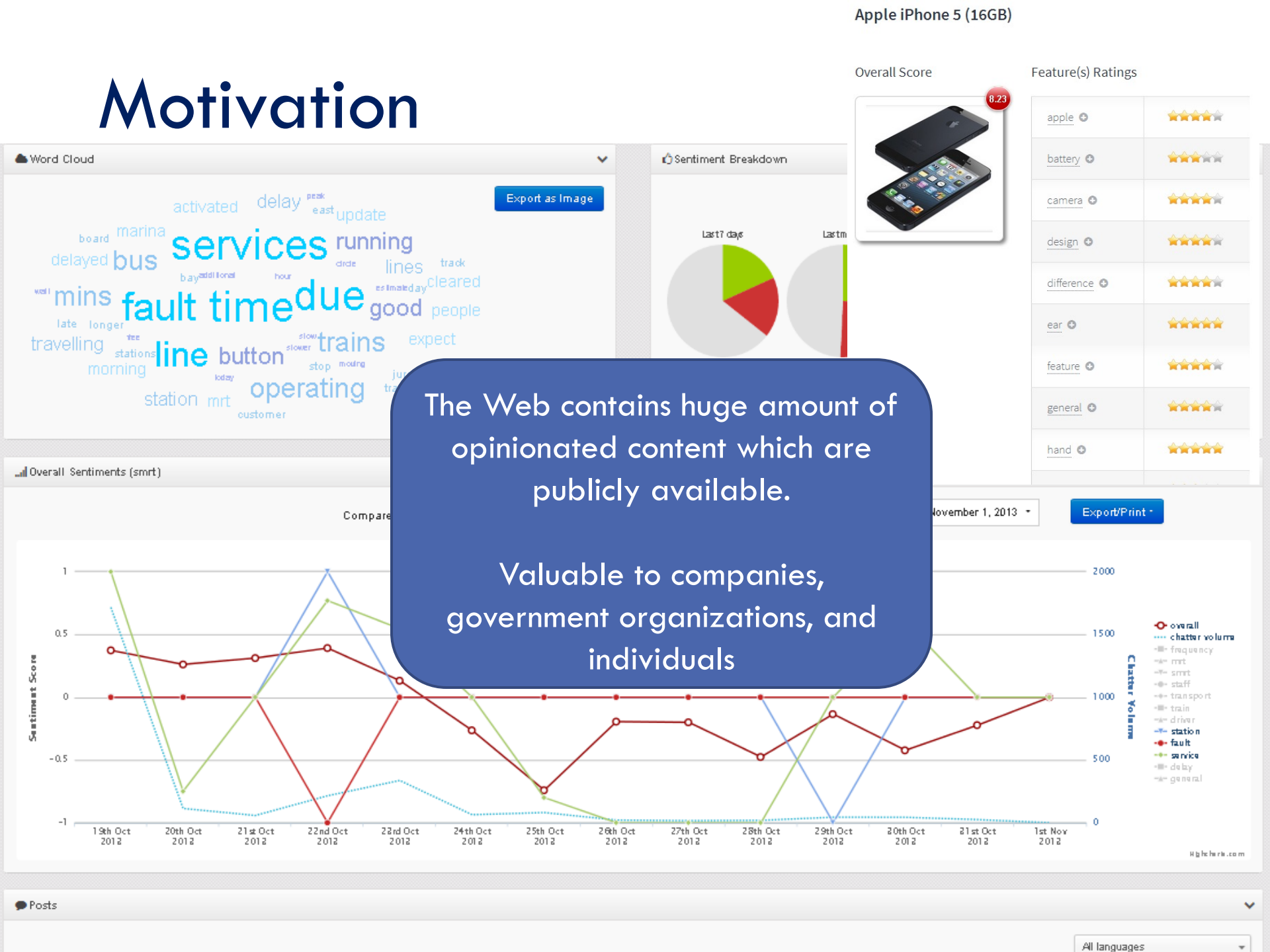
- Text Mining refers to Data Mining on Text (or Text Analytics)
  - Text ➡ Patterns ➡ Knowledge

Applications of Text Mining

- Information Extraction
- Information Retrieval
- Language Detection
- Language Translation
- Sentiment Analysis

Introduction to Sentiment Analysis

Motivation



The Web contains huge amount of opinionated content which are publicly available.

Valuable to companies, government organizations, and individuals

Sentiments

Great for running applications and choice of applications, runs them very smoothly with no glitches. Large screen and very clear images, touch screen very sensitive and responsive, very good at smooth zooming and image quality is excellent.

Battery life is limited will last about a day if not using too many apps. Signal seems to get lost easily, had to reboot once to regain signal.

What is Sentiment Analysis

- Task of extracting and analyzing **subjective content** (i.e. opinion)
  - To generalize how users feel about issues/things
- Instead of **sentiment analysis**, sometimes the term **opinion mining** is used
- Theoretically speaking, there are 2 subtasks:
  - **Subjectivity Classification**
  - **Sentiment Classification**
    (sentiment polarity : +ve/neu/-ve)

Examples of Opinions

Types of Opinions	Example
Product Reviews (Product)	Battery life is quite bad.
Forum Posting (General)	Tech coys dying. Easy money all drying up.
Customer Service Testimonials (Service)	The service has been spectacular.
Restaurant Reviews (Service)	Everything is perfect! The service, the food, the atmosphere.
Tweets (Issue/Policy)	Russia is waging a war against Ukraine. Going to war for a bad cause is a terrible thing. Hope the Russian will push back on their government and protest against this unjustified war now they are affected.

Opinionated Content vs Factual Content

Opinions	Facts
An opinion is a statement which may or may not be backed by facts. I.e. just views of an individual/group	A statement that can be proven true or false
Subjective	Objective
Contains **polarity** – positive, negative, neutral	Can also contain polarity – positive, negative, neutral

- Nowadays, nobody really cares whether sentiment analysis is applied to only opinions and **sentiment analysis** is commonly treated as **sentiment classification**
  - E.g. Applying sentiment analysis to financial news in order to determine market sentiment for predicting stock price movement

Application of Sentiment Analysis

- Tracking sentiment of a product/brand/company
- Capturing market trend
- Tracking sentiment of policies/events
- Government Intelligence
- …

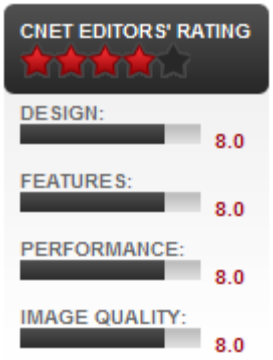
How To Analyze Content?

- Make use of clues to help in the analysis process
  - Non-text-based
    - Rating stars

  - Text-based
    - Analyze the linguistic properties of the content
    - Apply Natural Language Processing techniques

Non-text-based Clues

- Ratings
  - Overall
  - Individual feature rating



Kodak Playsport

Best budget pick

Kodak delivers decent HD video at a dirt cheap price for the point-and-shoot crowd.

Price: **$112.01 - $149.99** (check prices)

Excellent

Read full review

Text-based Clues

- Can consider under different levels
  - Word/Phrase level
  - Sentence level
  - Document level

Text-based Clues

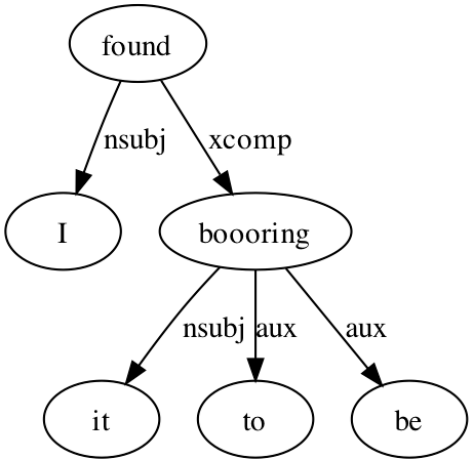
- Part of Speech (POS)
  - Noun
  - Verb
  - Adjective
  - Adverb
- **N-gram**
  - E.g.
    Large screen and very clear images
    far exceeded my expectation
    long lasting battery
- Word Frequencies

Large,screen,and
large_screen,screen_and,...
large_screen_and,
screen_and_very

stopwords, non-dictionary words: bugfested, lol

Syntactic Structure of Phrases/Sentences

- Dependency Parse
- Connectives
  - Conjunctions
    - X *and* Y, X *but* Y
  - Conditionals
    - *If, Unless, Even if, Only if, In case, etc*
  - Negation
    - *Not* X
  - Comparison
    - X *better than* Y, X *as good as* Y



Levels of Sentiment Analysis

- Sentiment analysis can be applied to different levels:
  - **Document-level**
  - **Sentence-level**
  - **Word-level**
  - **Aspect-level**

Document-level Sentiment Analysis

- When applied to the document-level, we aim to derive the sentiment for the entire document
  - This is the most coarse-grained
  - Writer of the document would most likely have differing opinions even within the document
  - So, knowing the overall sentiment is usually not that helpful

Sentence-level Sentiment Analysis

- Apply sentiment analysis to each sentence
  - Resulting in finer grain sentiment analysis
  - The assumption is that the sentiment within a sentence is fixed
    - Not always true
    - But works a lot better than document-level sentiment analysis
  - Could combine (by averaging) the sentiment of each sentence to obtain the overall sentiment of the document

Word-level Sentiment Analysis

- Sentiment analysis on the word-level is different from the rest in that it is more about determining the sentiment orientation of a word
- The final outcome is we get a list of word and its corresponding polarity (aka a **sentiment lexicon**)
- General approach:
  - Make use of a list of **seed positive/negative words** (which is manually created)
  - Expand the list of positive/negative words by comparing the unseen words with the seed positive/negative words
  - **Any idea how this is done using the knowledge you have obtained from the previous course?**

Issue with Coarse-grained Sentiment Analysis

Great for running applications and choice of applications, runs them very smoothly with no glitches. Large screen and very clear images, touch screen very sensitive and responsive, very good at smooth zooming and image quality is excellent.

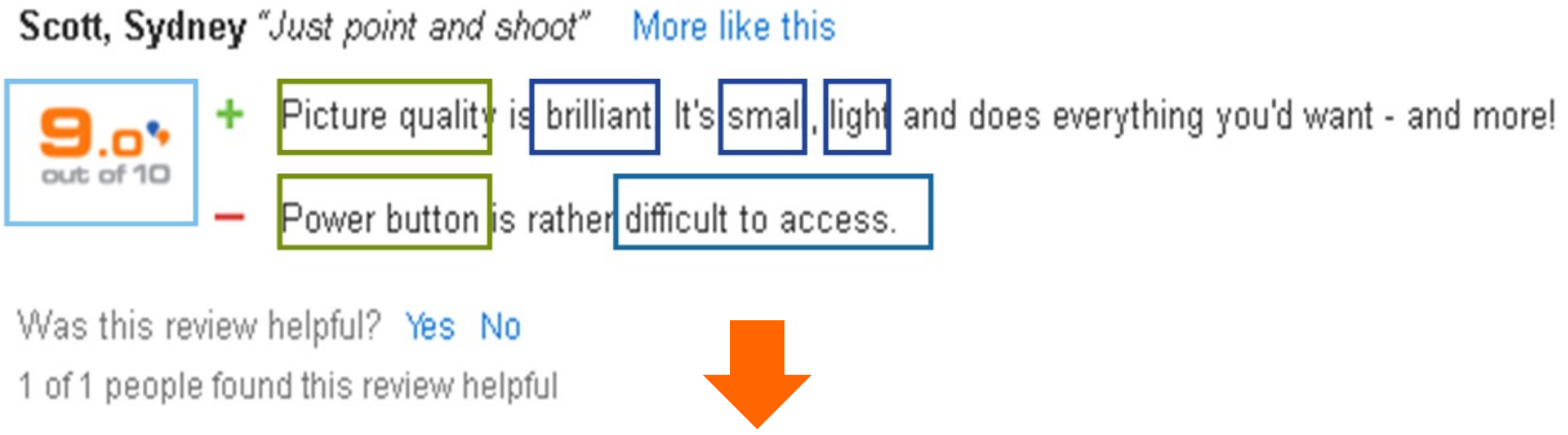
Battery life is limited will last about a day if not using too many apps. Signal seems to get lost easily, had to reboot once to regain signal.

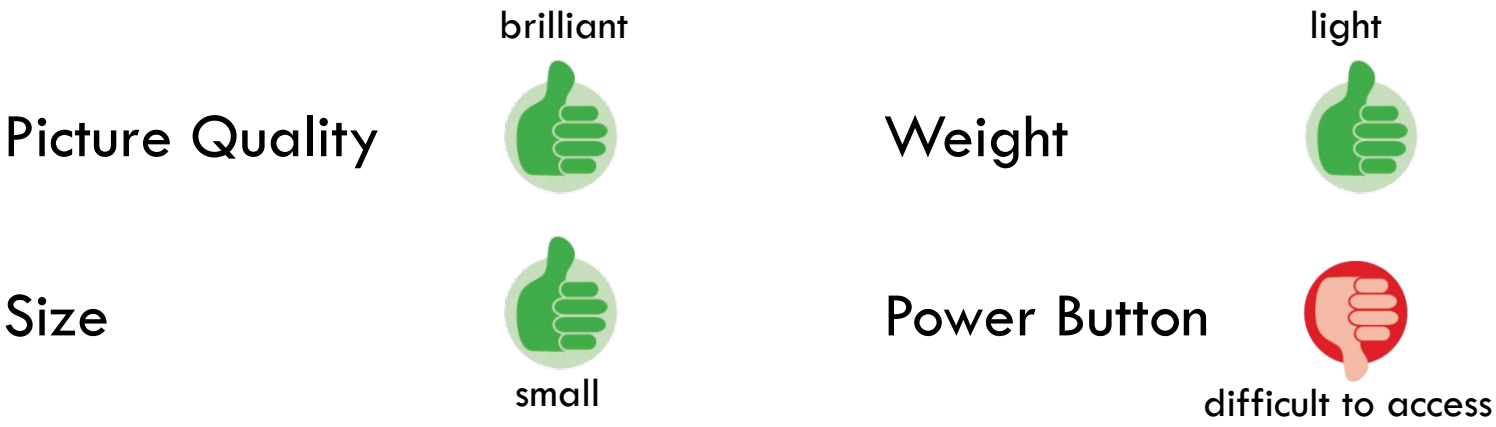
Often, a document or even a sentence can contain more than one sentiment and mixed sentiments

➔

Better to perform fine-grained sentiment analysis

Aspect-level Sentiment Analysis





Sentiment Analysis Approaches

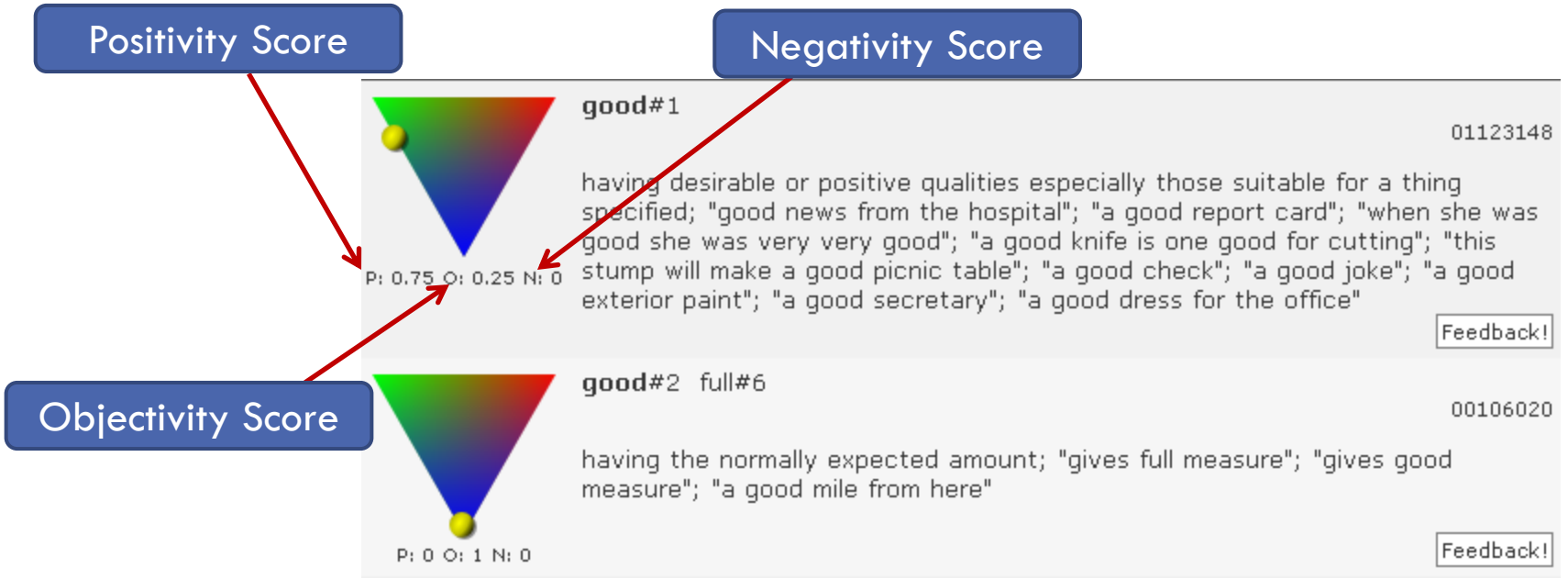
- Knowledge-based Approaches
  - Use of Lexicon
    - Using the help of a lexicon to determine the sentiment of the document, sentence, word, aspect
  - Rule-based
- Machine Learning Approaches
  - Text classification
    - Will cover this later

Sentiment Lexicons

- SentiWordNet
  - Based on WordNet 3.0
  - https://github.com/aesuli/SentiWordNet
- Subjectivity Lexicon
  - http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/
- Opinion Lexicon
  - www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar

SentiWordNet

- General-purpose sentiment lexicon
  - Each word in the lexicon is associated with a polarity
- Based on WordNet
- Each synset is associated with 3 numerical scores



Opinion Lexicon

Negative Words	Positive Words
2-faced	a+
2-faces	abound
abnormal	abounds
abolish	abundance
abominable	abundant
abominably	accessable
abominate	accessible
abomination	acclaim
abort	acclaimed
aborted	acclamation
aborts	accolade
…	…

Simple Lexicon-based Sentiment Analyzer

- Build a simple Lexicon-based sentiment analyzer by counting the number of positive and negative words
- Approach:
  - Tokenize the document/sentence
  - Check each token against a sentiment lexicon to see whether it is found in the lexicon
  - Initialize **counter** as 0
    - Increment the **counter** if found as a positive word
    - Decrement the **counter** if found as a negative word
  - **counter** = 0 ➔ neutral
    counter > 0 ➔ positive
    counter < 0 ➔ negative

Simple Lexicon-based Sentiment Analyzer

- Example:

Score	Text
-1	i believe in hate at first sight !
2	cool pics^) I love it
0	I HATE IT THAT I LOVE YOU
-1	UK is already in a recession, new data from Bank of England indicates

This method is commonly used in the finance research sector where the sentiment analysis is not so sophisticated.

No need to be super sophisticated also since sentiment analysis (of the news headlines) does not equal to the market sentiment

Most Sentiment Lexicon

good 

wonderful 

bad 

terrible 

Problem With Most Sentiment Lexicon

expensive

price

look

Polarity of Sentiment words depends on both the aspect and domains!

tiny

(computer monitor)
size

size
(portable battery charger)

Domain & Aspect-Specific Sentiment Lexicon

digital-cameras

Show 100 entries Search:

Word	Aspect	Polarity
zoom		
quality	zoom	+ve
decent	zoom	+ve
smooth	zoom	+ve
excellent	zoom	+ve
amazing	zoom	+ve
precise	zoom	+ve
superb	zoom	+ve
great	zoom	+ve
wonderful	zoom	+ve

SENTIMENT ANALYSIS FUNDAMENTALS

INTRODUCTION

DR LEK HSIANG HUI