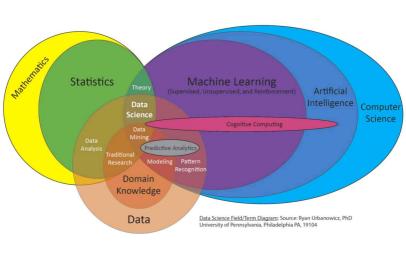
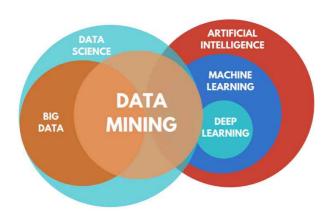
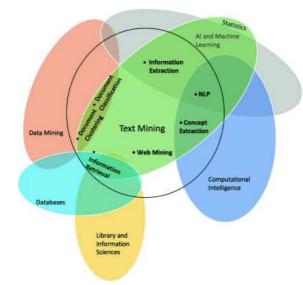
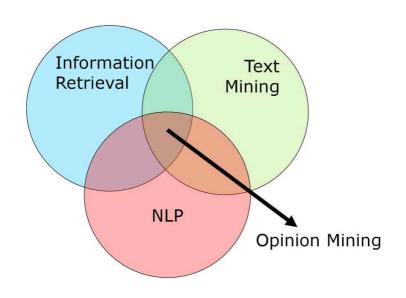
Sentiment Analysis

Lexicon based Sentiment Analysis on corpus and dictionary









Sentiment Analysis

Why sentiment analysis?

Positive or Negative

Unbelievably disappointing

Full of zany characters and richly applied satire, and some great plot twists

This is the greatest screwball comedy ever

It was pathetic. The worst part about it was the boxing scenes

살짝 지루한 감이 있었지만, 과격한 장면 없고 부드러워서 좋았어요. 아이들과 감동받고 갑니다.

초중반까지만진짜재밌게봤다...십자가로툭찔러서디질거면악령하지마라제발미리미리성불이나처하라고



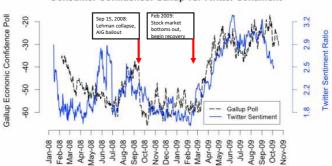
Shopping

쇼핑몰리뷰

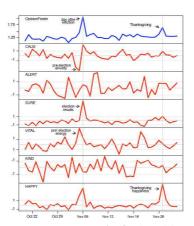


Consumer Confidence

Consumer Confidence: Gallup vs. Twitter Sentiment



Twitter Sentiment



Stock Market

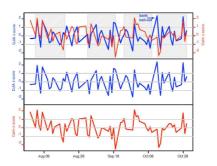
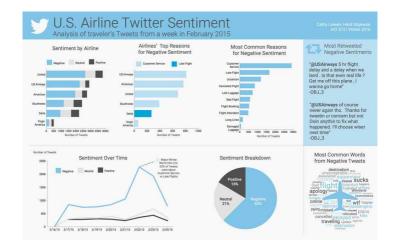


Fig. 3. A panel of three graphs. The top graph shows the overlap of the day-to-day difference of DIA values (blue: \mathbb{Z}_{D_t}) with the GPOMS' Calm time series (red: \mathbb{Z}_{X_t}) that has been lagged by 3 days. Where the two graphs overlap the Calm time series predict changes in the DIA closing values that occur 3 days later. Areas of significant congruence are marked by gray areas. The middle and bottom graphs show the separate DIIA and GPOMS' Calm time series.

Target Sentiment



Sentiment Analysis

Opinion extraction

Opinion mining

Sentiment mining

Subjectivity analysis

Why sentiment analysis?

Movie Is this review positive or negative?

Products What do people think about the new

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

What is sentiment?

Sentiment

Sentiment

Oxford A view or opinion that is held or expressed

Dictionaries

General **feeling** or **opinion**

A feeling or emotion

WordNet Sentiment (tender, romantic, or nostalgic feeling or emotion)

Opinion, sentiment, persuasion, view, thought

(a personal belief or judgment that is not founded on proof or certainty)

표준국어사전 이성에 대응되는 개념

외계의 대상을 오관으로 감각하고 지각하여 표상을 형성하는 인간의 인식 능력



Subjectivity

Synonym

Emotion A strong feeling **deriving from one**'s circumstances, mood, or relationships with others

Feeling An emotional state or reaction, strong emotion

Opinion A view or judgement formed about something, not necessarily based on fact or knowledge

Attitude A settled way of thinking or feeling about something



Subjective impression, not facts





Basic Emotions Ekman & Friesen (1975)





Happiness Sadness Fear



Anger Surprise Disgust



Noncognitive Motor Signs

- Prewired, Automatic
- Universal
 - Across Cultures
 - Across Species
- Blends of Emotion

I love her 🖊 I think she loves me

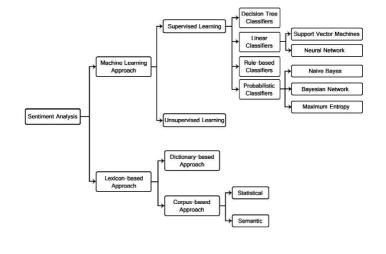
Sentiment Analysis

What is Sentiment Analysis?

Using NLP, statistics, or machine learning methods

to extract, identify, or otherwise characterize the sentiment content of a text





Detection of attitudes

```
Holder(Source) of attitude
Target(Aspect) of attitude
```

Type of attitude

```
Set of types (like, love, hate, etc.)
```

Weighted polarity (positive, negative, neutral, ... with strength)

Text containing the attitude

How to analyze a text?

Simplest

Is the attitude of this text positive or negative?

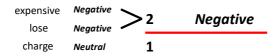
Complex

Rank the attitude of this text from 1 to 5 (start-rating)

Advanced

Detect the target, source, or complex attitude types





Objective

"The Force Awakens" is the highest grossing "Star Wars" movie.



Subjective

"The Empire Strikes Back" is the best "Star Wars" movie.



Subjectivity vs Objectivity



Importance of Objectivity



Sentiment strength

Domain-specific context

It was incredibly unpredictable
Positive? Negative?

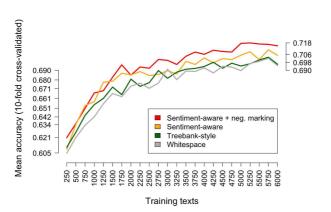
Different strength

It is such a nice phone, although a little large Positive? Negative?

Sentence	Sentiment	
This is good book.	Positive	
This is great book.	Very Positive	
This is a book.	Neutral	
This is not a great book.	Negative	
This is the worst book that I have ever come across.	Very Negative	

Negation

Train on OpenTable; test on 6000 IMDB reviews (1% = 60 reviews)



Sentiment Dictionary

Manual (Human based)

A significant amount of manual work and time is required

Dictionary based

A newly coined words, abbreviations, and figurative expressions are excluded

Corpus based

Finding opinion words with context specific orientations in a large corpus

How to build dictionary?

Semi-supervised learning

Use a small amount of information

A few labeled examples (seed words)

A few hand-built patterns (syntactic structure)

To bootstrap a lexicon

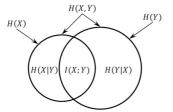
Using big data (lexicon resources, corpus)

Sentiment Measure

Mutual Information

Measure of the mutual dependence between the two random variables

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} P_{X,Y}(x,y) \log \frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}$$
$$P_X(x) = \sum_{Y} P_{X,Y}(x,y)$$



Pointwise Mutual Information

Measure of the mutual dependence

A pair of outcomes x and y belonging to discrete random variables X and Y

$$\begin{split} PMI(x;y) &= log \frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)} = log \frac{P_{X,Y}(x|y)}{P_X(x)} = log \frac{P_{X,Y}(y|x)}{P_Y(y)} \\ &-\infty \leq PMI(X;Y) \leq \min(-log P_X(x), -log P_Y(y)) \end{split}$$

X	у	p(x, y)
0	0	0.1
0	1	0.7
1	0	0.15
1	1	0.05

	p(x)	p(y)
0	0.8	0.25
1	0.2	0.75

$$\begin{split} PMI(0;0) &= \log_2 \frac{0.1}{0.8 \times 0.25} = -1 \\ PMI(0;1) &= \log_2 \frac{0.7}{0.8 \times 0.75} = 0.222 \\ PMI(1;0) &= \log_2 \frac{0.15}{0.2 \times 0.25} = 1.584 \\ PMI(1;1) &= \log_2 \frac{0.05}{0.2 \times 0.75} = -1.584 \end{split}$$

Exercise

word 1	word 2	count word 1	count word 2	count of co-occurrences	PMI	
puerto	ouerto rico 1938		1311	1159	10.0349081703	
hong	kong	2438	2694	2205	9.72831972408	
los	angeles	3501	2808	2791	9.56067615065	
carbon	dioxide	4265	1353	1032	9.09852946116	
prize	laureate	5131	1676	1210	8.85870710982	
san	francisco	5237	2477	1779	8.83305176711	
nobel	prize	4098	5131	2498	8.68948811416	
ice	hockey 5607		3002	1933	8.6555759741	
star	trek	8264	1594	1489	8.63974676575	
car	driver	5578	2749	1384	8.41470768304	
it	the	283891	3293296	3347	-1.72037278119	
are	of	234458	1761436	1019	-2.09254205335	
this	the	199882	3293296	1211	-2.38612756961	
is	of	565679	1761436	1562	-2.54614706831	
and	of	1375396	1761436	2949	-2.79911817902	
a	and	984442	1375396	1457	-2.92239510038	
in	and	1187652	1375396	1537	-3.05660070757	
to	and	1025659	1375396	1286	-3.08825363041	
to	in	1025659	1187652	1066	-3.12911348956	
of	and	1761436	1375396	1190	-3 70663100173	

1,159/50,000,952

PMI-IR

Find semantic orientation(SO)

Estimate co-occurrence probabilities using hit counts

$$SO(x) = PMI(x; positive) - PMI(x; negative)$$

$$= log \frac{P_{X,Y}(x, positive)}{P_X(x)P_Y(positive)} - log \frac{P_{X,Y}(x, negative)}{P_X(x)P_Y(negative)}$$

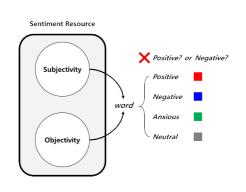
$$= log \frac{P_{X,Y}(x, positive)P_X(x)P_Y(negative)}{P_X(x)P_Y(positive)P_{X,Y}(x, negative)}$$

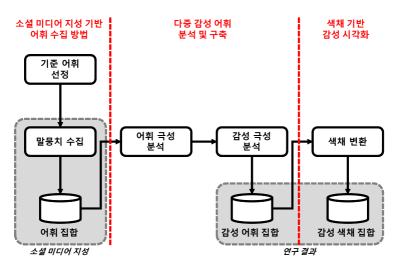
$$= log \frac{P_{X,Y}(x, positive)P_Y(negative)}{P_Y(positive)P_{X,Y}(x, negative)}$$

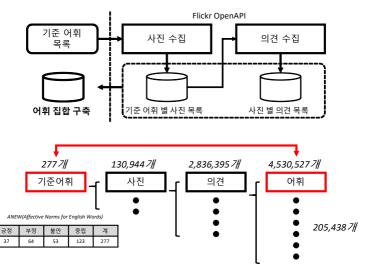
$$= log \frac{hits(x, positive) hits(negative)}{hits(positive)hits(x, negative)}$$

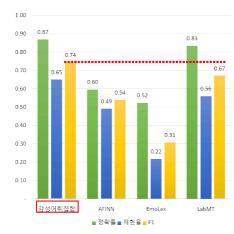
Experiments

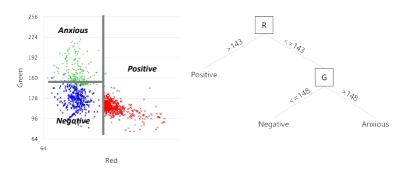
using Social Media(Flickr)











× Positive × Negative + Anxious

A single word



A multinomial word

surprised = (0.5668, -0.3609, **0.9355**, -1.1414)



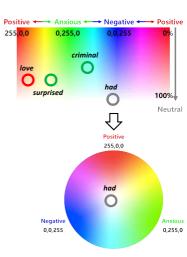
criminal = (-0.9877, 0.7714, **0.9204**, -0.7042)



A neutral word

had = (0.0177, 0.1122, 0.0101, -0.1401)





Twitter Set (Stanford) #3 by tpryan #4 by vcu451 #5 by chadfu #6 by SIX15 #7 by yamarama #8 by GeorgeVHulme #9 by Seth937 Positive @kenburbary You'll love your Kindle2. I've had mine for a few months and never looked back. The new big one is huge! No need for remorse! :)

#10 by dcostalis #11 by PJ_King

#13 by ipeb

#12 by mandanicole

#14 by kylesellers #15 by theviewfans #16 by MumsFP #17 by vincentx24x #18 by cameronwylie



단어	PMI	SO-PMI	단어	PMI	SO-PMI
	(Pos.)	(PosNeg.)	E-*1	(Neg.)	(NegPos.
happy	0.3969	0.5224	sad	0.2946	0.3866
happiness	0.3884	0.5858	despair	0.2635	0.4938
joy	0.3823	0.5784	grief	0.2605	0.5374
cheerful	0.3811	0.7223	tragic	0.2592	0.5860
faithful	0.3632	0.7814	solemn	0.2586	0.7233
carefree	0.3618	0.8727	tragedy	0.2585	0.5049
cheer	0.3564	0.6513	tears	0.2585	0.4105
enjoying	0.3558	0.6911	feeble	0.2584	0.4103
eager	0.3556	0.7624	100010	0.2584	0.7226
lively	0.3552	0.7640	bankrupt alienation		0.7226
devoted	0.3549	0.8590		0.2581	0.0.00
amazed	0.3547	0.7622	rejected	0.2581	0.6180
pleased	0.3546	0.7608	blame	0.2575	0.6247
glorious	0.3545	0.6263	penalty	0.2575	0.7231
vitality	0.3542	0.8570	ashamed	0.2575	0.6170
hopeful	0.3541	0.7032	weakness	0.2573	0.6173
stamina	0.3539	0.6986	pathetic	0.2573	0.5515
pleasant	0.3531	0.6932	dud	0.2572	0.5513
medal	0.3526	0.7229	quilt	0.2570	0.7232
bliss	0.3489	0.6430	cramp	0.2569	0.7203
excited	0.3469	0.6707	desperate	0.2565	0.5569
smile	0.3465	0.4645	punish	0.2564	0.5764
tender	0.3459	0.5645	burden	0.2563	0.5704
warmth	0.3453	0.5637	quilty	0.2560	0.5313
hug	0.3438	0.5269	awful	0.2557	0.5244
humerous	0.3433	0.8324			0.0
sparkle	0.3415	0.6021	curse	0.2557	0.5334
laugh	0.3393	0.4850	misery	0.2551	0.5360
joke	0.3354	0.4817	gloom	0.2546	0.5553
ardent	0.3347	0.8454	loss	0.2540	0.4985
friendly	0.3319	0.5939	failure	0.2538	0.5569
trust	0.3310	0.5080	terrible	0.2538	0.5809
proud	0.3240	0.4844	regret	0.2533	0.5182
peaceful	0.3147	0.4696	sadness	0.2525	0.4489