Sentiment Analysis on the Social Web

Dr. Nicole NOVIELLI

Assistant Professor

Dipartimento di Informatica

Università degli Studi di Bari 'Aldo Moro'

COLLAB, Collaborative Development Group

Three-day course @Eindhoven University of Technology on
Affect Detection in the Social Programmer Ecosystem

Outline - Day 1

- Introduction
- Part 1 - Theoretical Background (Affect Modeling)
- Part 2 - Mining Opinions and Emotions from Text

(15 mins) Break

- (75 mins) Part 3 - An Overview on Lexical Resources for Sentiment Analysis

Sentiment Analysis

- Also known as opinion mining, is the task of identifying the ***subjectivity*** (neutral vs. emotionally loaded) and the ***polarity*** (positive vs. negative semantic orientation) of a text, by exploiting natural language processing and computational linguistics.

Anger Fear Disgust Surprise Happiness Sadness

Foundations and Trends in Information Retrieval
Vol. 2, No 1-2 (2008) 1–135

DOI: xxxxxx

now
the essence of knowledge

Opinion mining and sentiment analysis

Bo Pang[1] and Lillian Lee[2]

[1] *Yahoo! Research, 701 First Ave. Sunnyvale, CA 94089, U.S.A., bopang@yahoo-inc.com*
[2] *Computer Science Department, Cornell University, Ithaca, NY 14853, U.S.A., llee@cs.cornell.edu*

Abstract

An important part of our information-gathering behavior has always been to find out what other people think. With the growing availability and popularity of opinion-rich resources such as online review sites and personal blogs, new opportunities and challenges arise as people now can, and do, actively use information technologies to seek out and understand the opinions of others. The sudden eruption of activity in the area of opinion mining and sentiment analysis, which deals with the computational treatment of opinion, sentiment, and subjectivity in text, has thus occurred at least in part as a direct response to the surge of interest in new systems that deal directly with opinions as a first-class object.


AFFECTIVE COMPUTING
Rosalind Picard

THEORETICAL BACKGROUND

Typology of Affective States

Duration ++

Intensity ++

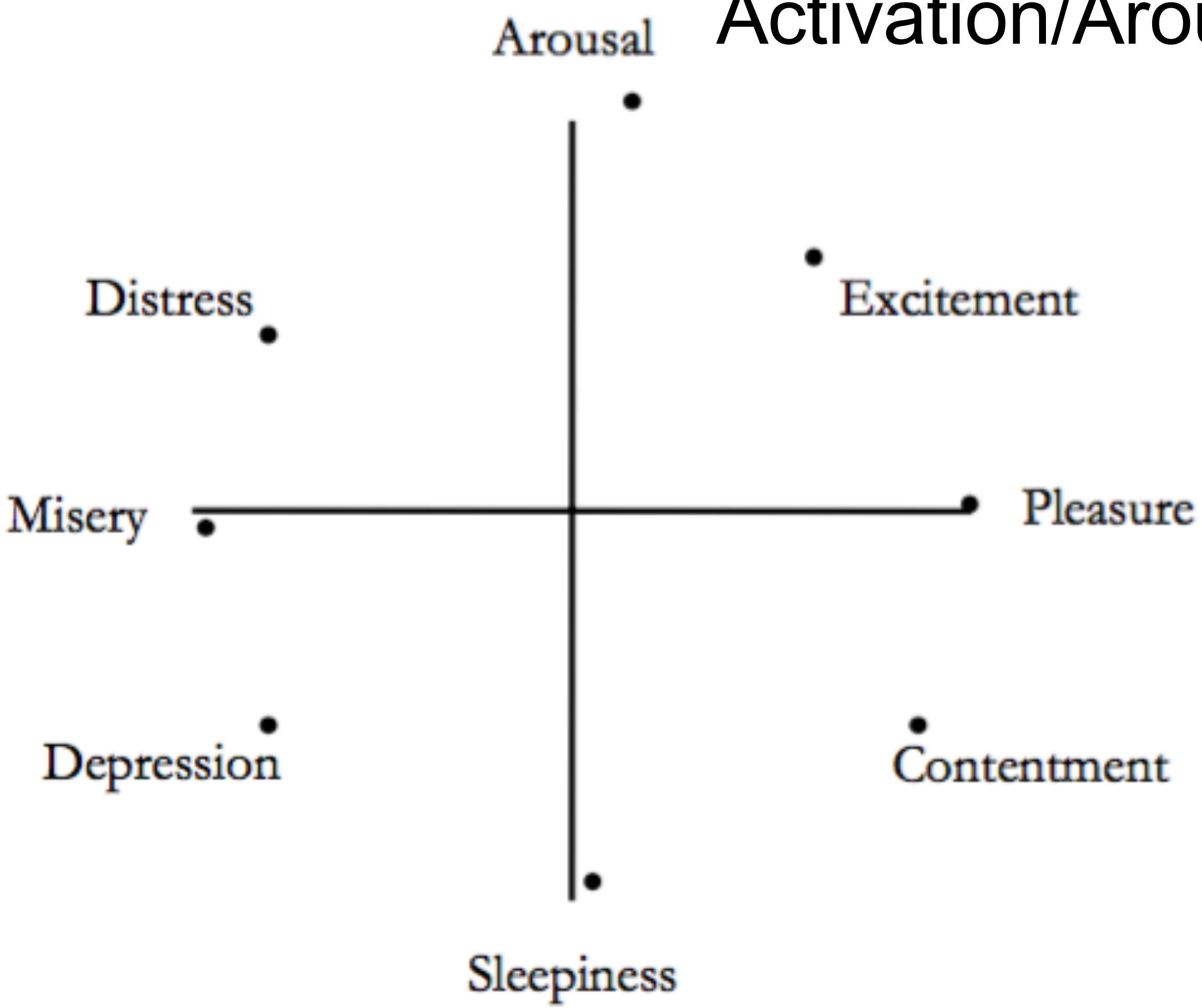
- **Personality traits**: stable personality dispositions and typical behavior tendencies
  - *nervous, anxious,reckless, morose, hostile, jealous*
- **Attitudes**: enduring, affectively coloured beliefs, dispositions towards objects or persons
  - *liking, loving, hating, valueing, desiring*
- **Interpersonal stances**: affective stance toward another person in a specific interaction
  - *friendly, flirtatious, distant, cold, warm, supportive, contemptuous*
- **Mood**: diffuse non-caused low-intensity long-duration change in subjective feeling
  - *cheerful, gloomy, irritable, listless, depressed, buoyant*
- **Emotion**: brief organically synchronized … evaluation of an major event as significant
  - *angry, sad, joyful, fearful, ashamed, proud, elated*

Scherer, 1984. Emotion as a Multicomponent Process: A model and some cross-cultural data. In P. Shaver, ed., Review of Personality and Social Psych 5: 37-63.

Continuous vs. Discrete Emotion Models (1/4)

Circumplex Model of Affect

Activation/Arousal



Polarity or Valence

Russel, 1980 - A circumplex model of affect. J.Personality and Social Psychology, 39, 1161-1178.

Continuous vs. Discrete Emotion Models (2/4)

Ekman's basic emotions

Anger Fear Disgust Surprise Happiness Sadness

Continuous vs. Discrete Emotion Models (3/4)

Lazarus' framework

Based on appraisal theory

Negative Emotions	Positive Emotions	Mixed Emotions
anger	happiness	hope
fright	pride	compassion
anxiety	relief	gratitude
guilt	love	
shame		
sadness		
envy		
jealousy		
disgust		

Affect and SA

Existing theories can be very useful when analysing data, particularly for:

- Formulating annotation schemes
- Understanding natural and derived labels and clusters
- Reducing the dimensionality of existing labels
- Finding useful sentiment contrasts and alignments

SENTIMENT ANALYSIS: TASKS AND APPROACHES

Message Polarity Classification

- The amount of user-generated content on the Internet has risen exponentially over the last decade
- A huge amount of documents is now available for researchers and practitioners
- Application domain: classifying reviews, microposts, etc.
- Goal: define the overall polarity of a text
  - i.e. positive, negative, neutral

Message Polarity Classification

1. I have *studied* all day but tomorrow I'm going out with *friends*! **:D** 
2. That’s **awful**. 
3. Most common nights to order pizza: NYE, Jan 1, day before Thanksgiving, Super Bowl Sunday, Halloween. 

Two Main Approaches

- Lexical approach (Unsupervised)
  - Exploits sentiment lexicons
- Supervised, based on machine learning
  - Training of supervised classifier on manually annotated dataset (gold standard)
  - Neural Networks, SVM, Naïve Bayes, …

Supervised Approach

1. Build a gold standard
   - Human-coded corpus for training, validation and testing
2. Extract features
   - Based on linguistic knowledge:
     - Linguistic devices to express emotions and opinions
     - Role of negations, intensifiers, etc.
     - Part of speech
   - Shallow linguistic features, e.g. n-gram
     - Token, lemma, stem
3. Choose machine learning algorithm
   - SVM, Convolutional Neural Networks, Naive Bayes, etc.
4. Validate on test-set
   - Ideally from another domain/source

Supervised Approach

1. Build a gold standard
   - Human-coded corpus for training, validation and testing
2. Extract features
   - Based on linguistic knowledge:
     - Linguistic devices to express emotions and opinions
     - Role of negations, intensifiers, etc.
     - Part of speech
   - Shallow linguistic features, e.g. n-gram
     - Token, lemma, stem
3. Choose machine learning algorithm
   - SVM, Convolutional Neural Networks, Naive Bayes, etc.
4. Validate on test-set
   - Ideally from another domain/source

Build a gold standard

- Manual annotation: define clear guidelines
  - Emotion framework as a theoretical ground
- Assess robustness and validity of the annotation schema
  - Interrater Agreement, e.g. using Kappa
- Define criteria to assign gold labels
  - e.g., using majority agreement
  - Community is now considering to use weighted schema for label assignment to deal with the high subjectivity of emotion annotation
- Problem: subjective and time consuming, not based on self report.

Supervised Approach

1. Build a gold standard
   - Human-coded corpus for training, validation and testing
2. Extract features
   - Based on linguistic knowledge:
     - Linguistic devices to express emotions and opinions
     - Role of negations, intensifiers, etc.
     - Part of speech
   - Shallow linguistic features, e.g. n-gram
     - Token, lemma, stem

Text preparation needed

3. Choose machine learning algorithm
   - SVM, Convolutional Neural Networks, Naive Bayes, etc.
4. Validate on test-set
   - Ideally from another domain/source

Text Preprocessing

- Tokenization

http://sentiment.christopherpotts.net/tokenizing.html

- Stemming or Lemmatization

http://sentiment.christopherpotts.net/stemming.html

- Advanced linguistic structure

http://sentiment.christopherpotts.net/lingstruc.html

Tokenization

- Splitting a string into its constituent parts
- The right algorithm depend on the application
- Special attention should be devoted to user-contributed texts on social media
  - >:-(
  - @SentimentSymp: can't wait for the #Sentiment talks!
  - YAAAAAAY!!! >;:-D http://sentimentsymposium.com/.

Tokenization

- Demo (also deals with negations) http://sentiment.christopherpotts.net/tokenizing/
- A basic, Twitter-aware tokenizer http://sentiment.christopherpotts.net/code-data/happyfuntokenizing.py
- Tools:
  - Twokenize: https://github.com/brendano/ark-tweet-nlp/blob/master/src/cmu/arktweetnlp/Twokenize.java
  - Stanford NLP toolkit: http://nlp.stanford.edu/software/

Stemming

- Method for clustering distinct word forms
  - Elimination/regularization of word suffixes
- Helps reducing the vocabulary size
- Especially effective for small data sets
- Problem: stemmers destroy sentiment distinctions
  - Inflected forms of the same word might convey different sentiment polarity or intensity
  - http://sentiment.christopherpotts.net/stemming.html

Effect of Stemming

Positiv	Negativ	Stemmed
captivation	captive	captiv
common	commoner	common
defend	defendant	defend
defense	defensive	defens
dependability	dependent	depend
dependable	dependent	depend
desirable	desire	desir
dominance	dominate	domin
dominance	domination	domin
extravagance	extravagant	extravag
home	homely	home
pass	passe	pass
patron	patronize	patron
prosecute	prosecution	prosecut
affection	affectation	affect
capitalize	capital	capit
closeness	close	close
commitment	commit	commit

- Porter Stemming on Harvard General Inquirer

Advanced Linguistic Structures

- Negations
- Part-of-speech (POS) tagging
- Dependency parsing

Negations

- Mild words behave like their opposites
  - bad ≈ not good
  - good ≈ not bad
- Strong words tend to neutral
  - not superb
  - not horrible

Handling Negations

- Append a _NEG suffix to words in the scope of negation
  - Appearing between a negation and a clause-level punctuation mark, i.e.: not comma
- I didn't enjoy it -> *I didn’t enjoy_neg it_neg*

Christoper Potts, Sentiment Symposium Tutorial - http://sentiment.christopherpotts.net/lingstruc.html

Sanjiv Das and Mike Chen. 2001. Yahoo! for Amazon: extracting market sentiment from stock message boards. In *Proceedings of the 8th Asia Pacific Finance Association Annual Conference*.

Pang, Bo; Lillian Lee; and Shivakumar Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. ACL.

POS Tagging

- Equal tokens may have different parts of speech
  - It is fine if you fine me, but don't rub it in!
  - *It/PRP is/VBZ **fine/JJ** if/IN you/PRP **fine/VBP** me/PRP ,/, but/CC don't/NN rub/VBP it/PRP in/IN !/.*
- Stanford Log-linear Part-Of-Speech Tagger
  - http://nlp.stanford.edu/software/tagger.shtml

http://sentiment.christopherpotts.net/lingstruc.html#tagged

Supervised Approach

1. Build a gold standard
   - Human-coded corpus for training, validation and testing
2. Extract features
   - Based on linguistic knowledge:
     - Linguistic devices to express emotions and opinions
     - Role of negations, intensifiers, etc.
     - Part of speech
   - Shallow linguistic features, e.g. n-gram
     - Token, lemma, stem
3. Choose machine learning algorithm
   - SVM, Convolutional Neural Networks, Naive Bayes, etc.
4. Validate on test-set
   - Ideally from another domain/source

Choose Machine Learning Algorithm

- Experiment with different classifiers
  - ***Parameter tuning*** using K-fold cross validation, usually with K=10
  - Train vs. Test, usually 70/30 split
- Assess performance[2]

http://sentiment.christopherpotts.net/classifiers.html#accuracy

  - Accuracy = # Correctly Classified Cases / # Cases
    - Inaccurate assessment of performance with highly imbalanced datasets
  - Precision = True Positive / All Positive Outcomes
  - Recall = True Positive / All Positive Documents
  - F-measure = (2·Precision·Recall) / (Precision+Recall)

[2] F. Sebastiani, "Machine learning in automated text categorization," ACM Comput. Surv., vol. 34, no. 1, pp. 1–47, Mar. 2002.

Choose Machine Learning Algorithm

- Weka (Java library with GUI)
  - http://www.cs.waikato.ac.nz/ml/weka/
  - Input: arff or csv
- Caret (R library)
  - http://caret.r-forge.r-project.org/
  - Input: csv

Supervised Approach

1. Build a gold standard
   - Human-coded corpus for training, validation and testing
2. Extract features
   - Based on linguistic knowledge:
     - Linguistic devices to express emotions and opinions
     - Role of negations, intensifiers, etc.
     - Part of speech
   - Shallow linguistic features, e.g. n-gram
     - Token, lemma, stem
3. Choose machine learning algorithm
   - SVM, Convolutional Neural Networks, Naive Bayes, etc.
4. Validate on test-set
   - Ideally from another domain/source

Limitations

- Technically optimized to a domain (i.e., having the highest accuracy scores)
- May use ***indirect*** indicators of sentiment and therefore give misleading results
  - Names of countries afflicted by war (Iran, Iraq) are typically associated with bad news
  - ‘Killer’ features for technology products become quickly obsolete
  - Proper names

- *Indirect* vs *Direct Affective Words*
  - Predictions based on machine learning might indicate topics in common rather than actual sentiment in common
  - This distinction is based on the OCC model and introduced by WordNet Affect

Lexical Approach

- Identifying the presence of terms from a lexicon of known sentiment-bearing words or phrases.
- Incorporate sentiment word lists (sentiment lexicons) into rule-based systems
- Sentiment lexicons can be manually or automatically created
  - E.g., the adjectives extracted from a set of texts in a gold standard
- Used in many types of sentiment analysis

We will be analyzing sentiment lexicons in a while…

Lexical Approach

- The lexicon method can be supplemented with other information
  - Emoticon lists
  - Semantic rules, i.e. for dealing with negation (Neviarouskaya, Prendinger, & Ishizuka, 2007; Taboada et al., 2011).
- Lexicons can be derived from a variety of sources
  - We will see many of them in a while…
- Specifically handling compound words

  (Neviarouskaya, Prendinger, & Ishizuka, 2011)
- To enhance performance, the seed emotion word list can be enriched with domain-specific terms
  - i.e. "small" being a general positive word for portable electronic device reviews (Yue Lu, Castellanos, Dayal, & Zhai, 2011; Velikovich, Blair-Goldensohn, Hannan, & McDonald, 2010).

We will see how Sentistrength works…

However…

- Nothing is as simple as it appears…

Message Polarity Classification is Far from Trivial

1. I have *studied* all day but tomorrow I'm going out with *friends*! **:D** 
2. I have *studied* all day but tomorrow I'm going out with *friends* 
3. Most common nights to order pizza: NYE, Jan 1, day before Thanksgiving, Super Bowl Sunday, Halloween. 
4. That's **awful**. 
5. 'That's **awful**', he said ?

+/-EffectWordNet: Sense-level Lexicon Acquisition for Opinion Inference

Yoonjung Choi and **Janyce Wiebe**
Department of Computer Science
University of Pittsburgh
yjchoi, wiebe@cs.pitt.edu

Abstract

Recently, work in NLP was initiated on a type of opinion inference that arises when opinions are expressed toward events which have positive or negative effects on entities (*+/-effect events*). This paper addresses methods for creating a lexicon of such events, to support such work on opinion inference. Due to significant sense ambiguity, our goal is to develop a sense-level rather than word-level lexicon. To maximize the effectiveness of different types of information, we combine a graph-based method using WordNet[1] relations and a standard classifier using

entity may be propagated to other entities via opinion inference rules. They give the following example:

(1) *The bill would curb skyrocketing health care costs.*

The writer expresses an explicit **negative** sentiment (by *skyrocketing*) toward the **object** (*health care costs*). The event, *curb*, has a **negative effect** on *costs*, since they are reduced. We can reason that the writer is **positive** toward the **event** because it has a negative effect on *costs*, toward which the writer is negative. From there, we can reason that the writer is **positive** toward *the bill*, since it is the agent of the positive event. Deng and Wiebe

Classifying the Overall Polarity of a Text

- Traditional approaches usually developed on reviews
- User-contributed text on social media require specific attention
  - Broken syntax (is POS tagging reliable?)
  - Space limitation and abbreviation (need to adapt text preparation and sentiment word lists, i.e. LOL)
  - Platform-specific features, i.e. user mentions, hashtags, etc. requiring ad hoc tokenization and conveying specific semantics

Subjectivity and Polarity Detection in Social Web

- Semeval has a dedicated track to Sentiment Analysis on Twitter *only* since 2013

  http://alt.qcri.org/semeval2016/index.php?id=tasks

- This year, best performance achieved with CNN for the subjectivity and polarity detection task

  http://alt.qcri.org/semeval2016/task4/

Aspect-Based Sentiment Analysis

http://alt.qcri.org/semeval2016/task5/

- Identification of the ***aspects*** of given target entities and the ***sentiment expressed*** for each aspect
- Summarization of the content of users' reviews

*"The **food** was lousy - too sweet or too salty and the **portions** tiny"*

- The sentiment with respect to these two aspects is negative

Stance Detection

- Automatically determining from text whether the author is ***in favor*** of the given target, ***against*** the given target, or whether ***neither inference*** is likely.

Target: *legalization of abortion*
Tweet: *A foetus has rights too! Make your voice heard.*

- Humans can deduce from the tweet that the speaker is likely against the target.

Stance Detection

- Relevant for many applications
  - Information retrieval
  - Text summarization
  - Textual entailment
    - directional relation between text fragments
    - "*t* entails *h*" ($t \Rightarrow h$) if, typically, a human reading *t* would infer that *h* is most likely true

Sentiment Quantification

- Given a set of tweets about a given topic, estimate the ***distribution*** of the tweets across the polarity classes
  - Also expressed as *n*-point scale

Barranquero, J., Diez, J., & del Coz, J. J. (2015). Quantification-oriented learning based on reliable classifiers. Pattern Recognition, 48(2), 591–604.

Esuli, A., & Sebastiani, F. (2015). Optimizing text quantifiers for multivariate loss functions. ACM Transactions on Knowledge Discovery and Data, 9 (4), Article 27. (see especially Section 2 for a brief survey on quantification methods) (**contains links to quantification software)**

SA of Irony and Figurative Language

- Polarity Classification (Semeval 2015)
  - Classification of polarity of tweets containing irony and metaphors
  - http://alt.qcri.org/semeval2015/task11/
- Irony detection and Polarity Classification (Evalita)
  - http://www.evalita.it/2016

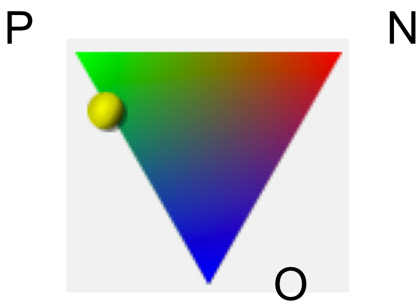
AN OVERVIEW ON LEXICAL RESOURCES FOR SENTIMENT ANALYSIS

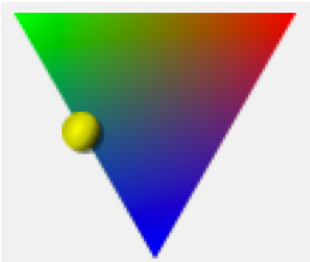
SentiWordNet 3.0

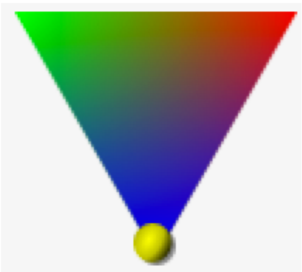
http://sentiwordnet.isti.cnr.it/

- Assigns to each synset of WordNet three sentiment scores: positivity, negativity, objectivity
- Sample code to approximate the sentiment of a word http://sentiwordnet.isti.cnr.it/code/SentiWordNetDemoCode.java

SentiWordNet







Scores	Sense ID and gloss
as an adjective	
Positive = 0.75 Negative = 0 Objective = 0.25	**good#1** - Having desirable or positive qualities especially those suitable for a thing specified (as in ‘a good joke’)
as a noun	
Positive = 0.5 Negative = 0 Objective = 0.5	**good#1** benefit (as in ‘for your own good’)
Positive = 0 Negative = 0 Objective = 1	**good#4** commodity, article of commerce

http://sentiwordnet.isti.cnr.it/
Esuli and Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. LREC 2006
Baccianella et al. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining, LREC 2010

SentiWordNet

Scores	Sense ID and gloss
as an adjective	
Positive = 0.75 Negative = 0 Objective = 0.25	**good#1** **Having desirable or positive qualities especially those suitable for a thing specified (as in ‘a good joke’)**
Positive = 0 Negative = 0 Objective = 1	**good#2** **having the normally expected amount (as in ‘gives good measure’)**
Positive = 1 Negative = 0 Objective = 0	**good#6** **agreeable or pleasing** **(as in ‘we all had a good time’)**
as a noun	
Positive = 0.5 Negative = 0 Objective = 0.5	**good#1** **benefit (as in ‘for your own good’)**
Positive = 0.875 Negative = 0 Objective = 0.125	**good#2** **moral excellence or admirableness (as in ‘there is good in people’)**
Positive = 0 Negative = 0 Objective = 1	**good#4** **commodity, article of commerce**

Bing Liu's Opinion Lexicon

- Positive words: 2006
- Negative words: 4783
- Useful properties: includes mis-spellings, morphological variants, slang, and social-media mark-up

https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html

MPQA

http://mpqa.cs.pitt.edu/

- Multi-Perspective Question Answering Subjectivity Lexicon
- Lexical clues manually compiled from several sources

Wiebe, Janyce; Theresa Wilson; and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation* 39: 165-210.

MPQA

http://mpqa.cs.pitt.edu/

- Multi-Perspective Question Answering Subjectivity Lexicon

	Strength	Length	Word	Part-of-speech	Stemmed	Polarity
1.	type=weaksubj	len=1	abandoned	pos1=adj	stemmed1=n	priorpolarity=negative
2.	type=weaksubj	len=1	abandonment	pos1=noun	stemmed1=n	priorpolarity=negative
3.	type=weaksubj	len=1	abandon	pos1=verb	stemmed1=y	priorpolarity=negative
4.	type=strongsubj	len=1	abase	pos1=verb	stemmed1=y	priorpolarity=negative
5.	type=strongsubj	len=1	abasement	pos1=anypos	stemmed1=y	priorpolarity=negative

Wiebe, Janyce; Theresa Wilson; and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation* 39: 165-210.

Harvard General Inquirer

http://www.wjh.harvard.edu/~inquirer/

- Attaches syntactic, semantic, and pragmatic information to POS tagged words

	Entry	Positiv	Negativ	Othtags	Defined
1	A			DET ART	...
2	ABANDON		Negativ	SUPV	
3	ABANDONMENT		Negativ	Noun	
4	ABATE		Negativ	SUPV	
5	ABATEMENT			Noun	
...					
35	ABSENT#1		Negativ	Modif	
36	ABSENT#2			SUPV	
...					
11788	ZONE			Noun	

Stone, Philip J; Dexter C. Dunphry; Marshall S. Smith; and Daniel M. Ogilvie. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. Cambridge, MA: MIT Press.

NRC Emotion Lexicon

- List of words associated to both polarity and specific emotion labels
  - 14,000 English words
- Created by exploiting crowdsourced annotations
  - i.e Mechanical Turk to annotate emotions invoked by common words and phrases
- http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

Saif M. Mohammad and Peter D. Turney. 2010. Emotions evoked by common words and phrases: using mechanical turk to create an emotion lexicon. *NAACL HLT 2010, ACL*

Saif Mohammad and Peter Turney, 2013. Crowdsourcing a Word-Emotion Association Lexicon, *Computational Intelligence,* 29 (3), 436-465, 2013.

NRC Emotion Lexicon

associate	positive:1	negative:0	anger:0	anticipation:0	disgust:0	fear:0	joy:0	sadness:0	surprise:0	trust:1
association	positive:0	negative:0	anger:0	anticipation:0	disgust:0	fear:0	joy:0	sadness:0	surprise:0	trust:1
assorted	positive:0	negative:0	anger:0	anticipation:0	disgust:0	fear:0	joy:0	sadness:0	surprise:0	trust:0
assortment	positive:0	negative:0	anger:0	anticipation:0	disgust:0	fear:0	joy:0	sadness:0	surprise:0	trust:0
assuage	positive:1	negative:0	anger:0	anticipation:0	disgust:0	fear:0	joy:0	sadness:0	surprise:0	trust:0
assumed	positive:0	negative:0	anger:0	anticipation:0	disgust:0	fear:0	joy:0	sadness:0	surprise:0	trust:0
assuming	positive:0	negative:0	anger:0	anticipation:0	disgust:0	fear:0	joy:0	sadness:0	surprise:0	trust:0
assumption	positive:0	negative:0	anger:0	anticipation:0	disgust:0	fear:0	joy:0	sadness:0	surprise:0	trust:0
assurance	positive:1	negative:0	anger:0	anticipation:0	disgust:0	fear:0	joy:0	sadness:0	surprise:0	trust:1
assure	positive:0	negative:0	anger:0	anticipation:0	disgust:0	fear:0	joy:0	sadness:0	surprise:0	trust:1
assured	positive:1	negative:0	anger:0	anticipation:0	disgust:0	fear:0	joy:0	sadness:0	surprise:0	trust:1
assuredly	positive:0	negative:0	anger:0	anticipation:0	disgust:0	fear:0	joy:0	sadness:0	surprise:0	trust:1
asterisk	positive:0	negative:0	anger:0	anticipation:0	disgust:0	fear:0	joy:0	sadness:0	surprise:0	trust:0
asteroids	positive:0	negative:0	anger:0	anticipation:0	disgust:0	fear:0	joy:0	sadness:0	surprise:0	trust:0
astigmatism	positive:0	negative:0	anger:0	anticipation:0	disgust:0	fear:0	joy:0	sadness:0	surprise:0	trust:0
astonishingly	positive:1	negative:0	anger:0	anticipation:0	disgust:0	fear:0	joy:0	sadness:0	surprise:1	trust:0

NRC Hashtag

- List of words and their associations with eight emotions
  - anger, fear, anticipation, trust,surprise, sadness, joy, and disgust
- The associations automatically computed from tweets with emotion-word hashtags
  - e.g., #happy and #anger.
- http://www.purl.com/net/lexicons

#Emotional Tweets, Saif M. Mohammad, In Proceedings of the First JointConference on Lexical and Computational Semantics (*Sem), June 2012,Montreal, Canada.

Using Hashtags to Capture Fine Emotion Categories from Tweets,Saif M. Mohammad and Svetlana Kiritchenko, To Appear in Computational Intelligence.

NRC Sentiment 140

- Automatically created from the Sentiment140 emoticon corpus of 1.6 million tweets
  - Using PMI as association metrics
- Entries for uni-, bi-grams, word pairs
  - unigrams-pmilexicon.txt: 62,468 terms
  - bigrams-pmilexicon.txt: 677,698 terms
  - pairs-pmilexicon.txt: 480,010 terms
- http://help.sentiment140.com/for-students

Mohammad, S., Kiritchenko, S., and Xiaodan, Z. 2013 NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets, SemEval-2013.

Wordnet Affect

A-label	Example of Synsets
EMOTION	noun ‘anger’, verb ‘fear’
MOOD	noun ‘animosity’, adjective ‘fear’
TRAIT	noun ‘aggressiveness’, adj. ‘competitive’
COGNITIVE State	noun ‘confusion’, adj. ‘dazed’
PHYSICAL State	noun ‘illness’
HEDONIC Signal	noun ‘hurt’, noun ‘suffering’
Emotion-eliciting SITUATION	noun ‘awkwardness’
Emotional RESPONSE	noun ‘cold sweat’, verb ‘tremble’
BEHAVIOR	noun ‘offense’, adj. ‘inhibited’
ATTITUDE	noun ‘intolerance’, noun ‘defensive’
SENSATION	noun ‘coldness’, noun ‘feel’

http://www.slideshare.net/metavivo/wordnet-affect071116

Strapparava and Valitutti. WordNet-affect: an affective extension of WordNet. LREC 2004

Building a Lexicon

- Hand-build lists of words or phrases
  - Evaluation of experts/trained human coders
  - Labels assigned based on agreement among coders
- Semi-automatic/bootstrapping procedures
  - Define a set of seeds
  - Enrich the initial list based on word association metrics (e.g., PMI, LSA-based cosine similarity, etc.)

Turney, Peter D. and Michael L. Littman. 2003. Measuring praise and criticism: inference of semantic orientation from association. *ACM Transactions on Information Systems* 21: 315-346.

Passaro et al., 2015. Item: a VectorSpace Model to Bootstrap an Italian Emotive Lexicon. CliC-it 2015.

Sentiment Tutorial Demo and Scripts

- **Access Demo** http://sentiment.christopherpotts.net/lexicons.html
- **Tokenization** http://sentiment.christopherpotts.net/tokenizing/results/
- **Use lexicons to score texts** http://sentiment.christopherpotts.net/textscores/
- (Lovely to meet you @RosieFortescue . Who looked beautiful as always. @bluebird) -> different lexicons cause different scores for texts
- **Practicum: Use lexicons to score words** http://sentiment.christopherpotts.net/lexicon/results/
- https://github.com/nnovielli/SAcourseMaterial