THE UNIVERSITY of EDINBURGH
informatics

Affect in Text

Using natural language programming to extract attitudes, opinion and sentiment from text.

Clare Llewellyn

With a lot of help from:
Liu, Bing. "Sentiment analysis and opinion mining." Synthesis Lectures on Human Language Technologies 5.1 (2012)

http://www.cs.uic.edu/~liub/FBS/sentiment-analysis-tutorial-2012.pdf

THE UNIVERSITY *of* EDINBURGH
informatics

Summary

- 1. What affect in text is and why study it

- 2. Levels of affect analysis
    - Subjective or objective
    - Opinion
    - Emotion

- 3. Levels of textual analysis
    - Document Level
    - Sentence Level
    - Entity Level

What is affect in text?

Identifying different points of view, identifying different emotive dimensions, and classifying text by opinion.

Why study affect in text?

Analysing affect in text means automatic identification and classification.

Allows us to add structure to unstructured text.

Allows us to deal with a larger volume of data.

A warning...

This is a hard problem. Humans infer emotions, generally through a combination of several factors. In text a lot of these factors are missing and even humans don’t always get this right.

Where extracted affect used:

- Subjective / objective text identification
  - Assisting in information retrieval
  - Question answering
  - Citation analysis / suggestion

- Sentiment analysis / opinion mining
  - Politics (public opinion, vote prediction)
  - Brand analysis (twitter complaints, customer reviews, ad location)
  - Medical analysis (depression, relationships)

- Emotion analysis
  - Affective state and learning

General Overview

Types of affect we may want to analyse:

1) Subjectivity
2) Opinion
3) Emotion

1) Subjectivity

Subjective or Objective: Classifying if something is a fact or not.

Objective: Sometimes the sun shines in Edinburgh

Subjective: I like that the sun shines in Edinburgh

Objective sentences can imply opinion

It has rained in Edinburgh for the past 17 days.

THE UNIVERSITY of EDINBURGH
informatics

2) Opinion

Opinion: Classing text as positive or negative, adding a value to that class.

Positive:

I like the weather in Edinburgh

Very positive:

I really like the sun shining in Edinburgh

Negative:

I don't like the winter weather in Edinburgh

Very negative:

I hate the constant freezing cold wind that blows through Edinburgh

2) Opinion - Complexities

Sometime what people mean isn't that straight forward:

Different types of opinion:

Explicit - *Cats are really nice.*

Implicit – *Cats don't need as much walking as dogs.*

Comparative – *Cats are much nicer than dogs.*

Sarcasm or irony – A cutting expression or remark, often expressing a contrasting opinion to their own or what is expected.

I really like the rain.

3) Emotion

Emotion: People's subjective feelings and thoughts.

Classifying text with emotional categories such as love, joy, surprise, anger, sadness, and fear.

I am happy that the sun is shining today.

2. Levels of text we might analyse

- Document Level
- Sentence Level
- Entity Level

Levels of Analysis – Document Level

Discover the overall opinion within a document or a general overview of many documents.

- Widely studied problem, especially in reviews
- You know the entity so the task is to identifying sentiment
- Many reviews come with an overall rating / number of stars

Levels of Analysis – Document Level

"Absolutley awful"

Reviewed 10 August 2009

Stayed here one night but had to check out the following day due to the shabby state of the hotel and the damp smell that hits you as soon as you walk in the front door. Everything about the hotel is tired and has a broken down feel to it. This has to be one of the the worst hotels I have ever stayed in and cannot believe that it has a 4 star rating as 2 star would be more appropriate.

Stayed August 2009, travelled with family

Value
Location
Rooms
Cleanliness
Service

Less

THE UNIVERSITY *of* EDINBURGH
informatics

Levels of Analysis – Document Level

"Lovely old style hotel"

Reviewed 1 week ago

y wife and I just returned from 3 night stay at this hotel.We book it in the Spring and Autumn and have never been disappointed.And once again they came up trumphs.The weather unfortunately was poor.However they have a games room with indoor bowling,table tennis ,snooker & pool.The hotel also in the Autumn has its own curling rinks and they have a restaurant above the curling where you can eat and watch the games.They have a lovely warn swimming pool and hot tub.So really the weather did not bother us.The rooms are clean and comfortable and the food in the restaurant is superb...........Roll on the Spring again.

Room Tip: Have never had a poor room.They have a lift If you have mobility problems.

See more room tips

Stayed October 2014, travelled as a couple

Location

Rooms

Service

Levels of Analysis – Document Level

Pang, et al (2002)

Looked at positive and negative movie reviews from IMDB.

Use the number of stars as and indicator of positivity / negativity

Supervised machine learning techniques to learn positive and negative models (neutral removed).

Features: **unigrams**, bigrams, POS, adjectives

Results: ~80% success rate, unigrams difficult to beat

Levels of Analysis – Document Level

Thwarted Expectation

I hate the Spice Girls. [3 things the author hates about them]. Why I saw this movie is a really, really, really long story, but I did, and one would think I'd despise every minute of it. But... Okay, I'm really ashamed of it, but I enjoyed it. I mean, I admit it's a really awful movie ...the ninth door of hell...The plot is such a mess that it's terrible. But I loved it.

(Pang, et al 2002)

Levels of Analysis – Sentence Level

- Simple sentences are easy:

I hate the Spice Girls.

I enjoyed the movie.

This has to be one of the the worst hotels I have ever stayed in and cannot believe that it has a 4 star rating as 2 star would be more appropriate.

- Identify the noun in the sentence. (Part Of Speech (POS) tagging)
- Look up the other words in a Sentiment Lexicon or train a classifier to differentiate between positive and negative (and neutral)
- Apply a sentiment score for the sentence
- Many commercial companies do sentiment analysis only slightly beyond this level.

Sentiment lexicon

Sentiment Lexicon: Dictionary of opinion words

How to build a sentiment lexicon:

- Collect words (dictionaries, literary and newspaper texts)
- Identify semantic concepts
- Extract affective dimensions from each word

Issues:

- The sets of words extracted extend beyond just sentiment of words

For example: The word anger refers to an emotion, animosity to a mood, and confusion to a cognitive state

- There are different senses of the same word.

For example, the word surprise may refer to a feeling “the astonishment you feel when something totally unexpected happens to you”, to an event “a sudden unexpected event”, or to an action “the act of surprising someone”.

Example Lexicon – LIWC

URL: http://www.liwc.net/tryonline.php

Linguistic Inquiry and Word Count was created to analyse groups of words group of words from basic emotional and cognitive dimensions.

1) Word Collection and semantic identification - Words were collected from each category PANAS, Roget's Thesaurus, and standard English dictionaries.

2) Rating - They then used judges to evaluate if the words were assigned correctly.

4,500 words in 76 categories

Category	Examples
Negate	aint, ain't, arent, aren't, cannot, cant, can't, couldnt, ...
Swear	arse, arsehole*, arses, ass, asses, asshole*, bastard*, ...
Social	acquainta*, admit, admits, admitted, admitting, adult, adults, advice, advis*
Affect	abandon*, abuse*, abusi*, accept, accepta*, accepted, accepting, accepts, ache*
Posemo	accept, accepta*, accepted, accepting, accepts, active*, admir*, ador*, advantag*
Negemo	abandon*, abuse*, abusi*, ache*, aching, advers*, afraid, aggravat*, aggress*,
Anx	afraid, alarm*, anguish*, anxi*, apprehens*, asham*, aversi*, avoid*, awkward*
Anger	jealous*, jerk, jerked, jerks, kill*, liar*, lied, lies, lous*, ludicrous*, lying, mad

(http://sentiment.christopherpotts.net/lexicons.html)

Example Lexicon - WordNet-Affect

URL: http://wndomains.fbk.eu/wnaffect.html

1) Word Collection and identification of semantic concepts

2) Rating: Core terms manually annotated with affective dimensions. Then this was propagated through the relevant WordNet synsets.
http://wordnet.princeton.edu/

	#Nouns	#Adjectives	#Verbs	#Adverbs	#Total
#Synsets	763	1462	322	327	2874
#Words	1285	2293	657	552	4787

Table 2: Affective synsets and words, grouped by part of speech

A-Labels	Examples
EMOTION	noun anger#1, verb fear#1
MOOD	noun animosisy#1, adjective amiable#1
TRAIT	noun aggressiveness#1, adjective competitive#1
COGNITIVE STATE	noun confusion#2, adjective dazed#2
PHYSICAL STATE	noun illness#1, adjective all_in#1
EDONIC SIGNAL	noun hurt#3, noun suffering#4
EMOTION-ELICITING SITUATION	noun awkwardness#3, adjective out_of_danger#1
EMOTIONAL RESPONSE	noun cold_sweat#1, verb tremble#2
BEHAVIOUR	noun offense#1, adjective inhibited#1
ATTITUDE	noun intolerance#1, noun defensive#1
SENSATION	noun coldness#1, verb feel#3

Table 4: A-Labels and corresponding example synsets

(Strapparava and Valitutti 2004)

Example Lexicon - SentiWordNet

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

1) Word Collection: Extends worknet synsets by assigning how positive, negative or objective a term is.

2) Rating: Semi Supervised Machine Learning: Built using multiple classifiers which are trained on different training sets. The score is defined as a proportion of how many classifiers define that term within a certain class.

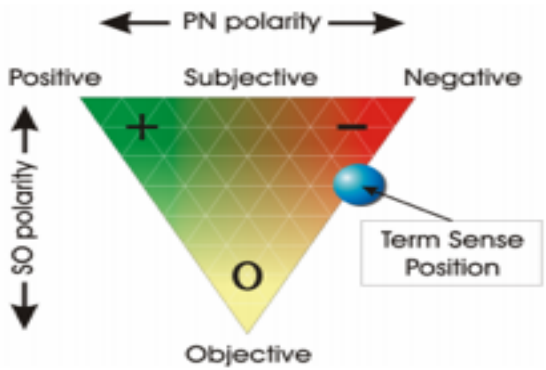


Figure 1: The graphical representation adopted by SENTI-WORDNET for representing the opinion-related properties of a term sense.

(Esuli and Sebastiani 2006)

THE UNIVERSITY of EDINBURGH
informatics

Example - Classifiers

Subjectivity classifier

Adjectives (preposterous), verbs, and certain phrases (dealt a blow) are statistically associated with subjective language.

(1) Classifiers learn to identify subjective and objective sentences

(2) These sentences are then used as training set to automatically learn extraction patterns associated with subjectivity

(3) Bootstrapping – sentences extracted in (2) are used to expand the training set

SYNTACTIC FORM	EXAMPLE PATTERN
<subj> passive-verb	<subj> was satisfied
<subj> active-verb	<subj> complained
<subj> active-verb dobj	<subj> dealt blow
<subj> verb infinitive	<subj> appear to be
<subj> aux noun	<subj> has position
active-verb <dobj>	endorsed <dobj>
infinitive <dobj>	to condemn <dobj>
verb infinitive <dobj>	get to know <dobj>
noun aux <dobj>	fact is <dobj>
noun prep <np>	opinion on <np>
active-verb prep <np>	agrees with <np>
passive-verb prep <np>	was worried about <np>
infinitive prep <np>	to resort to <np>

Figure 2: Syntactic Templates and Examples of Patterns that were Learned

(Riloff and Wiebe 2003)

Classifier or Lexicon?

Lexicons:

Lexicons are expensive and time consuming to produce

Not all lexicons are the same

They don't cover multiple languages

Machine Learning:

Generally require training data

Only covers the data that you have

They are hard to interpret and extend

Can use a combination of both

Level of Analysis - Entity Level

Opinion (entity, aspect, sentiment, opinion holder, time)

At the moment, artemisinin-based therapies are considered the best treatment, but cost about $10 per dose - far too much for impoverished communities.

— Seattle Times (Feb 16, 2012)

Entity Level

Opinion (entity, aspect, sentiment, opinion holder, time)

At the moment, artemisinin-based therapies are considered the best treatment, but cost about $10 per dose - far too much for impoverished communities.

— Seattle Times (Feb 16, 2012)

Entity Level

Opinion (entity, aspect, sentiment, opinion holder, time)

At the moment, artemisinin-based therapies are considered the best treatment, but cost about $10 per dose - far too much for impoverished communities.

— Seattle Times (Feb 16, 2012)

Entity Level

Opinion (entity, aspect, sentiment, opinion holder, time)

At the moment, artemisinin-based therapies are considered the best treatment, but cost about $10 per dose - far too much for impoverished communities.

— Seattle Times (Feb 16, 2012)

Entity Level

Opinion (entity, aspect, sentiment, opinion holder, time)

At the moment, artemisinin-based therapies are considered the best treatment, but cost about $10 per dose - far too much for impoverished communities.

— Seattle Times (Feb 16, 2012)

Example - Entity Level Sentiment Analysis

Popescu and Etzioni 2007

- Identify product features
- Identify opinions of those product features
- Determine polarity
- Rank opinions based on strength

Constraints:

a) conjunctions and disjunctions in the review text (and, but, or, yet)

b) rule templates

c) automatically derived morphological relationships (wonderful and wonderfully)

d) Use WordNet synonymy, antonymy, IS-A and morphological relationships between words

Table 2.11. Lexical Patterns Used to Derive Opinions' Relative Strength.

a, (*) *even b*	*a*, (*) *not b*
a, (*) *virtually b*	*a*, (*) *almost b*
a, (*) *near b*	*a*, (*) *close to b*
a, (*) *quite b*	*a*, (*) *mostly b*

THE UNIVERSITY of EDINBURGH
informatics

Example - Emotional Analysis in Learning

Studies have found that students learn well when they experience some confusion.

Students learn less well when they experience boredom

AutoTutor and Intelligent Tutoring System has pedagogical strategies to deal with students affective state and improve learning.

Used Supervised Learning on textual interaction between Autotutor and students.

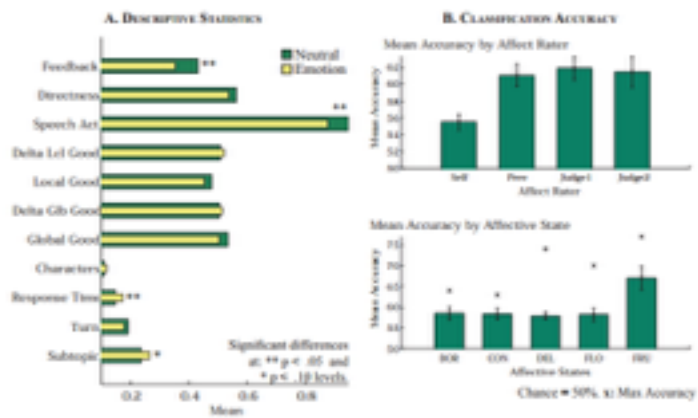


Figure 3: Descriptive statistics and reliability in affect detection from dialogue. For visualization purposes all features in 3a have been normalized to a 0-1 range.

(D'Mello et al 2007)

THE UNIVERSITY of EDINBURGH
informatics

Example - Sentiment Analysis in Social Media

High volume

Lack of context

Text specificities of medium:

- acronyms
- initialisms
- emoticons
- slang

Short segments of text

Straight to the point

Social Media Example -Vader

Hutto and Gilbert 2014

Uses a lexicon from the domain to create word lists which are specific to this social media text

Performs nearly as well as humans

1) Word collection and sentiment orientation from lexicons (integrating various lexicon and add in social media specific information)

2) Rating - Develop rules to identify intensity using patterns learned from human rating from Mechanical Turk

Social Media Example -Vader

Identify whole text factors by asking annotators to rate whole tweets come up with 5 rules:

1. Punctuation, namely the exclamation point (!), increases the magnitude of the intensity without modifying the semantic orientation.

   The food here is good!!!

2. Capitalization, specifically using ALL-CAPS to emphasize a sentiment-relevant word in the presence of other non-capitalized words, increases the magnitude of the sentiment intensity without affecting the semantic orientation.

   The food here is GREAT!

3. Degree modifiers (also called intensifiers, booster words, or degree adverbs) impact sentiment intensity by either increasing or decreasing the intensity.

   The service here is extremely good

4. The contrastive conjunction "but" signals a shift in sentiment polarity, with the sentiment of the text following the conjunction being dominant.

   The food here is great, but the service is horrible

5. By examining the tri-gram preceding a sentiment-laden lexical feature, we catch nearly 90% of cases where negation flips the polarity of the text.

   The food here isn't really all that great.

(Hutto and Gilbert 2014)

Social Media Example UK Connections

March:
#putin in the running to win nobel peace prize http://t.co/kce3geonly #lgbti

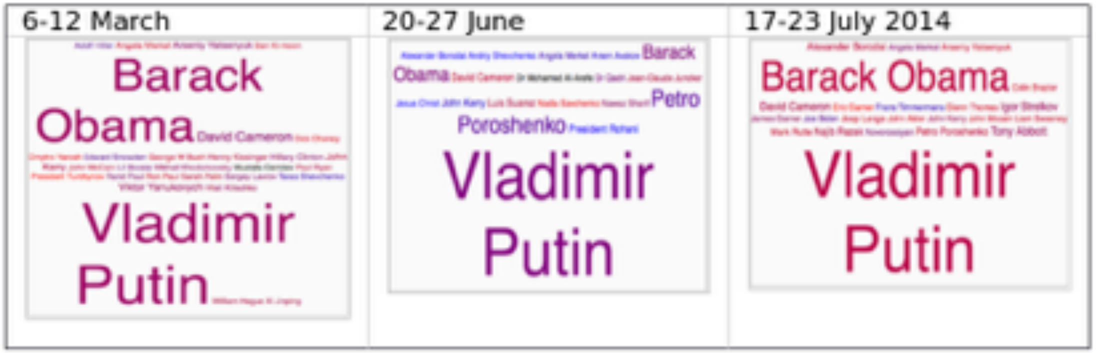
obama and putin: liars poker and the dangers of war obama has certainly lied about the ukraine crisis.

June:
Putin supports Ukraine peace plan: Russian President Vladimir Putin has said he supports a peace plan tabled b... http://t.co/nvFvaAnU1C

RT @EastOfBrussels: #Russia terrorists threaten to murder 100s of miners in #Ukraine if they don't support #Putin http://t.co/HQ2X4kP15e #g…

July:
@rndrive just so pleased we have 2 such trustworthy & reliable chaps as putin & abbott *on* this situation great comfort.

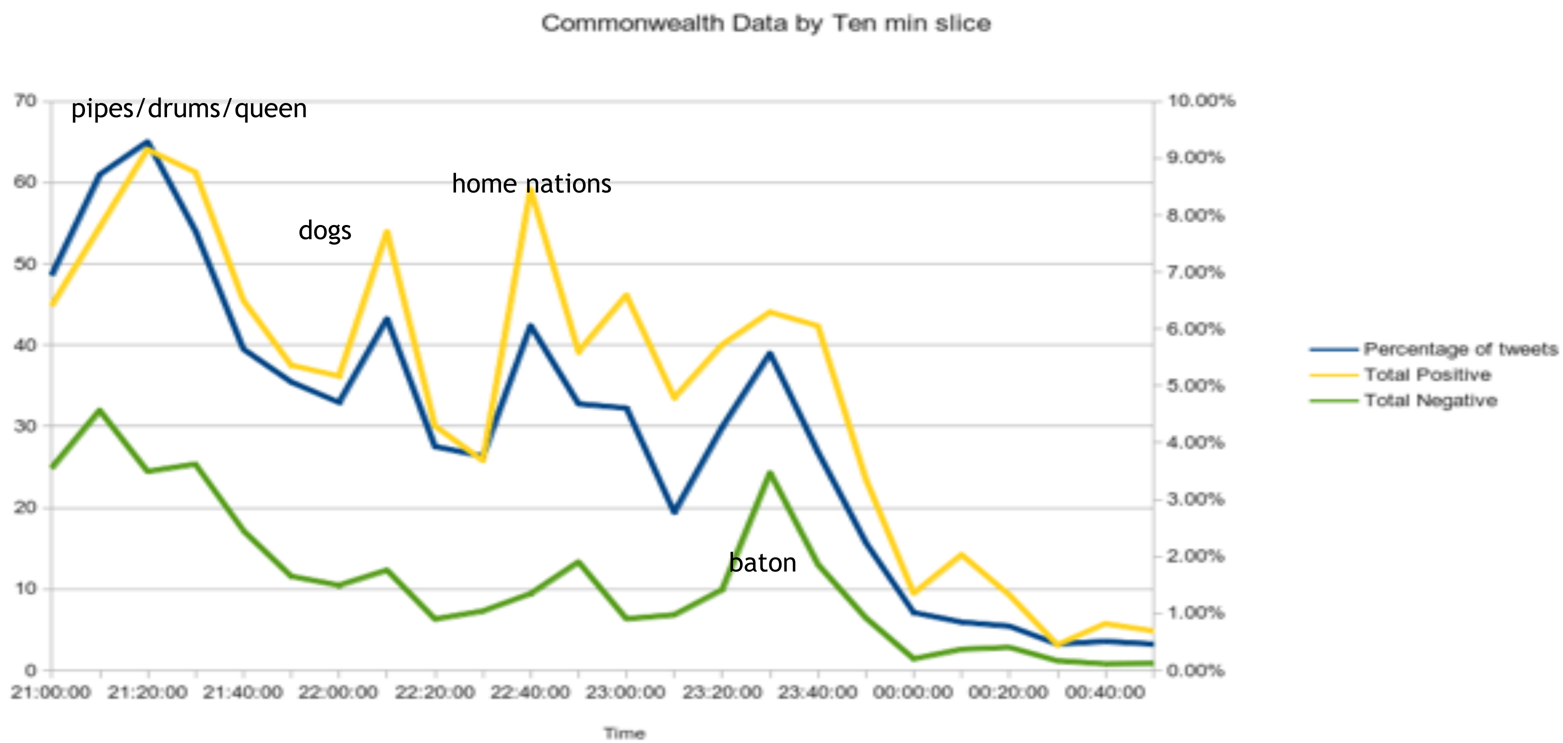
stupid ass murderer of his own people & piece of shit dictator #putin & his backwards ass country #russia is guilty of murder ! fuck #russia



THE UNIVERSITY *of* EDINBURGH
informatics

Social Media Example -UK Connections

Social Media Example -UKConnections



THE UNIVERSITY of EDINBURGH
informatics

Summary

General Overview – subjectiveness, opinion and emotion

Liu, Bing. "Sentiment analysis and opinion mining." Synthesis Lectures on Human Language Technologies 5.1 (2012)

Levels of Analysis

Document Level

Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up?: sentiment classification using machine learning techniques." *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*. Association for Computational Linguistics, 2002.

Sentence Level

C. Strapparava, A. Valitutti, WordNet-Affect: an affective extension of WordNet, Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC-2004) (2004), pp. 1083–1086 (Lisbon, Portugal)

Esuli and Sebastiani, SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining, Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC-2006)

Riloff and Wiebe, Learning extraction patterns for subjective expressions, Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (2003)

Entity Level

Popescu, Ana-Maria, and Orena Etzioni. "Extracting product features and opinions from reviews." Natural language processing and text mining. Springer London, 2007. 9-28.

Examples

D'Mello, Sidney, Rosalind Picard, and Arthur Graesser. "Towards an affect-sensitive autotutor." *IEEE Intelligent Systems* 22.4 (2007): 53-61.

Hutto, C. J., and Eric Gilbert. "Vader: A parsimonious rule-based model for sentiment analysis of social media text." Eighth International AAAI Conference on Weblogs and Social Media. 2014.