

COM6115: Text Processing

Sentiment Analysis: Approaches

Chenghua Lin

Department of Computer Science
University of Sheffield

By the end of the SA sessions, you will be able to:

- Explain the relevance of the topic
- Differentiate between objective and subjective texts
- List the main elements in a sentiment analysis system
- Provide a critical summary of the main approaches for the problem
- Explain how sentiment analysis systems are evaluated.

- Definition of the problem of sentiment analysis
- **Approaches to sentiment analysis**
- Evaluation of sentiment analysis approaches

Based on survey and slides by Bing Liu (University of Illinois at Chicago), 2012.

Two approaches to SA

- Lexicon-based
 - ◊ **Binary**
 - ◊ Gradable
- Corpus-based/Supervised machine learning

A simple approach to SA: lexicon-based

Use a lexicon of opinion/emotion words, like: good, bad, horrible, great, etc.

Rule-based sentiment classifier (sentence/document-level)

- 1 Rule-based **subjectivity classifier**: a sentence/document is **subjective** if it has at least n (say 2) words from the emotion words lexicon; a sentence/document is **objective** otherwise.
- 2 Rule-based **sentiment classifier**: for subjective sentences/documents, count positive and negative words/phrases in the sentence/document. If more negative than positive words/phrases, then **negative**; otherwise, **positive** (if equal, neutral).

Rule-based sentiment classifier (feature-level)

- Assume features can be identified in previous step by information extraction techniques, e.g., battery, phone, screen.
- For each feature, count positive and negative emotion words/phrases from the lexicon.
- If more negative than positive words/phrases, then negative; otherwise, positive (if equal, neutral).

Rule-based sentiment classifier (feature-based)

- Simple approach:
 - ◇ **Input**: a pair (f, s) , where f is a product feature and s is a sentence that contains f .
 - ◇ **Output**: whether the emotion on f in s is positive, negative, or neutral.
 - ◇ **Step 1**: work on the sentence s containing f .
 - ◇ **Step 2**: select emotion words in s : w_1, \dots, w_n .
 - ◇ **Step 3**: assign orientations for these emotion words: 1 = positive, -1 = negative, 0 = neutral.
 - ◇ **Step 4**: sum up the orientation and assign the orientation to (f, s) accordingly.
- More advanced approaches split the sentence in parts, e.g., based on BUT words (“but”, “except that”, ...).

Caveats

- Certain words have context-independent orientations, e.g. “great”.
- Other emotion words have **context-dependent** orientations, e.g.
 - ◇ **small** power consumption = positive
 - ◇ **small** screen = negative
 - ◇ **consume** valuable resources = negative
 - ◇ **consume** disgusting waste = positive
- One has to deal with negation, e.g.:
 - ◇ **not great** = negative
 - ◇ **not bad** = positive
- One has to deal with intensifiers:
 - ◇ **very good** is more positive than **good**
 - ◇ **extremely boring** = is more negative than **boring** or **very boring**

Can store more fine-grained sentiment information in lexicon and add additional **rules**.

Two approaches to SA

- Lexicon-based
 - ◊ Binary
 - ◊ **Gradable**
- Corpus-based/Supevised machine learning

Use of **ranges of sentiment** instead of a binary system, to deal with intensifiers like:

- absolutely, utterly, completely, totally, nearly, virtually, essentially, mainly, almost, e.g.: **absolutely awful**

And grading adverbs like:

- Very, little, dreadfully, extremely, fairly, hugely, immensely, intensely, rather, reasonably, slightly, unusually, e.g.: **a little bit cold**

Lexicon-based approach to SA - gradable

Rule-based gradable sentiment classifier

- Classifies general valence of a text (document-, sentence- or feature-level) based on the level of emotional content
- Level of emotional content given by:
 - The **lexicon**: word-lists with pre-assigned emotional weights, e.g:
Neg. dimension (C_{neg}): $\{-5, \dots, -1\}$, Pos. dimension (C_{pos}): $\{+1, \dots, +5\}$

bore	-3	careful	3
boring	-3	careless	-2
bother	-1	cares	2
brave	3	caring	3
bright	2	casual	2
brilliant	2	casually	2
broke	-1	certain	2
brutal	-3	challeng	2
burden	-1	champ	2
calm	2	charit	2
care	2	charm	2
cared	2	cheat	-3
carefree	2		

Lexicon-based approach to SA - gradable

- Ctd:

2 Additional **general rules** to change the original weights:

Negation rule: E.g.: “I am not good today”.

Emotion(good) = +3; “not” is detected in neighbourhood (of 5 words around); so emotional valence of “good” is decreased by 1 and sign is inverted → **Emotion(good) = -2**

Capitalization rule: E.g. “I am GOOD today”.

Emotion(good) = +3; Add +1 to positive words → **Emotion(GOOD) = +4**

Likewise, in “I am AWFUL today”.

Emotion(awful) = -4; Add -1 to negative words → **Emotion(awful) = -5**

Intensifier rule:

- Needs a list of intensifiers: “definitely”, “very”, “extremely”, etc.
- Each intensifiers has a weight, e.g. $\text{Weight}(\text{very})=1$;
 $\text{Weight}(\text{extremely})=2$
- The weight is added to positive terms
- The weight is subtracted from negative terms
- E.g.: “I am feeling very good”.
 $\text{Emotion}(\text{good})=+3$; emotional valence of “good” increased by 1
→ $\text{Emotion}(\text{good})=+4$
- E.g. “This was an extremely boring game”
 $\text{Emotion}(\text{boring})=-3$; emotional valence of “boring” decreased by -2
→ $\text{Emotion}(\text{boring})=-5$

Diminisher rule:

- Need a list: “somewhat”, “barely”, “rarely”, etc.
- Each intensifiers has a weight
- The weight is subtracted from positive terms
- The weight is added to negative terms
- E.g.: “I am somewhat good”.
 $\text{Emotion}(\text{good}) = +3$; emotional valence of “good” decreased by 1
 $\rightarrow \text{Emotion}(\text{good}) = +2$
- E.g. “This was a slightly boring game”
 $\text{Emotion}(\text{boring}) = -3$; emotional valence of “boring” increased by 1
 $\rightarrow \text{Emotion}(\text{boring}) = -2$

Lexicon-based approach to SA - gradable

Exclamation rule: Functions like intensifiers. E.g.: “Great show!!!”.

Emotion(great) = +3; Weight(!!!) = 2

→ Emotion(great) = 5

Emoticon rule: Each has its own emotional weight, like an emotion word.

E.g.: Emotion(😊) = +2; Emotion(☹) = -2. E.g.: “I can’t believe this product ☹”

→ Emotion(☹) = -2

Lexicon-based approach to SA - gradable

- Final decision based on ALL emotion words:
 - ◇ If $|C_{pos}| > |C_{neg}|$ then {positive}
 - ◇ If $|C_{pos}| < |C_{neg}|$ then {negative}
 - ◇ If $|C_{pos}| = |C_{neg}|$ then {neutral}
- E.g.: “He is brilliant but boring”:
Emotion(brilliant) = 2; Emotion(boring) = -3
→ $C_{pos} = 2$, $C_{neg} = -3$, so {negative}
- E.g.: “I am not good today”:
Emotion(good) = -2
→ $C_{pos} = 0$, $C_{neg} = -2$, so {negative}
- E.g.: “I am not GOOD today”: (Emotion(good)=3) → ???
- E.g.: “I am so surprised by this product!!! ☹”: (Emotion(☹)=-2) → ???

Lexicon-based approach to SA

Advantages:

- Works effectively with different texts: forums, blogs, etc.
- Language independent - as long as an up-to-date lexicon of emotion words is available
- Doesn't require data for training
- Can be extended with additional lexica, e.g. for new emotion words/symbols as they become popular, esp. in social media

Disadvantages:

- Requires a lexicon of emotion words, which should be fairly comprehensive, covering new words, abbreviations (LOL, m8, etc.), misspelled words, etc.

E.g.: In a dataset from MySpace, 95% of comments contained at least one spelling error!

Lexica of emotion words/phrases

For both binary and gradable approaches, how to obtain lexica of emotion words?

Task: Collect relevant words/phrases that can be used to express sentiment. Determine the **emotion** of these subjective word/phrases.

- **Manually:** word lists with pre-assigned emotional weights
- Semi-automatically
 - ◇ **Dictionary-based:** find synonyms/antonyms of seed emotion words in dictionaries like WordNet
 - ◇ **Corpus-based:** find synonyms/antonyms of seed emotion words in corpora

Lexica of emotion words/phrases (ctd)

Mostly **adjectives**

- **Positive**: e.g.: honest, important, mature, large, patient, ...
- **Negative**: harmful, hypocritical, inefficient, insecure

Verbs

- **Positive**: praise, love
- **Negative**: blame, criticize

Nouns

- **Positive**: pleasure, enjoyment
- **Negative**: pain, criticism

Phrases (esp. for collocations, but also alternative to having intensifiers weighted separately)

- **Positive**: high intelligence, low cost
- **Negative**: little variation, many problems

Lexica of emotion words/phrases (ctd)

Semi-automatically created resources, such as:

- **SentiWordNet**: Wordnet is a database with words grouped into sets of synonyms (synsets), and organised by semantic relations between them: synonyms, antonyms, hypernyms, etc. SentiWordNet is a version of it with one of three sentiment scores for each synset: **positivity**, **negativity**, **objectivity**.


Manually created resources, such as:

- **Linguistic Inquiry and Word Count (LIWC) lexicon**: made by psychologists with lists of words with various **emotional and other dimensions**.
- **General Inquirer**: terms with various types of **positive** or **negative** semantic orientation.




Lexica of emotion words/phrases (ctd)

SentiWordNet

sentiwordnet.isi.cnr.it/search.php?q=terrific

 **SentiWordNet**

ADJECTIVE

 P: 0.25 C: 0.5 N: 0.25	terrific#1 very great or intense; "a terrific noise"; "a terrific thunderstorm storm"; "fought a terrific battle"	01513619
Feedback on SentiWordNet values: <input type="button" value="They are OK"/> <input type="button" value="Suggest your values."/>		
 P: 0.75 C: 0.25 N: 0	wondrous#1 wonderful#1 tremendous#2 terrific#2 rattling#1 marvelous#1 marvellous#1 howling#1 grand#4 fantastic#2 extraordinarily good or great ; used especially as intensifiers; "a fantastic trip to the Orient"; "the film was fantastic!"; "a howling success"; "a marvelous collection of rare books"; "had a rattling conversation about politics"; "a tremendous achievement"	01676517
Feedback on SentiWordNet values: <input type="button" value="They are OK"/> <input type="button" value="Suggest your values."/>		
 P: 0 C: 0.375 N: 0.625	terrifying#1 terrific#3 causing extreme terror; "a terrifying wall"	00196449
Feedback on SentiWordNet values: <input type="button" value="They are OK"/> <input type="button" value="Suggest your values."/>		

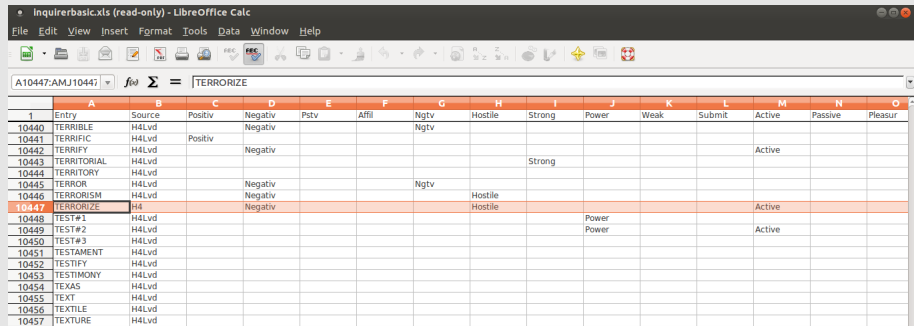
Lexica of emotion words/phrases (ctd)

Linguistic Inquiry and Word Count lexicon

Category	Abbrev	Examples	Words In Category
Psychological Processes			
Social processes	social	Mate, talk, they, child	455
Family	family	Daughter, husband, aunt	64
Friends	friend	Buddy, friend, neighbor	37
Humans	human	Adult, baby, boy	61
Affective processes	affect	Happy, cried, abandon	915
Positive emotion	posemo	Love, nice, sweet	406
Negative emotion	negemo	Hurt, ugly, nasty	499
Anxiety	anx	Worried, fearful, nervous	91
Anger	anger	Hate, kill, annoyed	184
Sadness	sad	Crying, grief, sad	101
Cognitive processes	cogmech	cause, know, ought	730
Insight	insight	think, know, consider	195

Lexica of emotion words/phrases (ctd)

General Inquirer: words classified in many categories, including: positive (1,915) and negative (2,291).



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Entry	Source	Positiv	Negativ	Pstv	Affil	Ngvtv	Hostile	Strong	Power	Weak	Submit	Active	Passive	Pleasur
10440	TERRIBLE	H4Lvd	Positiv	Negativ			Ngvtv								
10441	TERRIFIC	H4Lvd													
10442	TERRIFY	H4Lvd		Negativ									Active		
10443	TERRITORIAL	H4Lvd							Strong						
10444	TERRITORY	H4Lvd													
10445	TERROR	H4Lvd		Negativ			Ngvtv								
10446	TERRORISM	H4Lvd		Negativ				Hostile							
10447	TERRORIZE	H4		Negativ				Hostile					Active		
10448	TEST#1	H4Lvd													
10449	TEST#2	H4Lvd								Power			Active		
10450	TEST#3	H4Lvd								Power					
10451	TESTAMENT	H4Lvd													
10452	TESTIFY	H4Lvd													
10453	TESTIMONY	H4Lvd													
10454	TEXAS	H4Lvd													
10455	TEXT	H4Lvd													
10456	TEXTILE	H4Lvd													
10457	TEXTURE	H4Lvd													

Free dictionary:

<http://www.wjh.harvard.edu/~inquirer/homecat.htm>

Lexica of emotion words/phrases (ctd)

Semi-automatically created from seed words: start with **seed positive and negative words**:

- Search for synonyms/antonyms in **dictionaries** like WordNet; OR
- Build **patterns** from seed words/phrases to search on large **corpora**, like the Web:
 - ◇ “beautiful and” (+)
 - ◇ “low cost but” (-)
 - ◇ “very nice and ” (+)

Lexica of emotion words/phrases (ctd) - from corpora

Google "very nice and"

Web Images Maps Shopping More Search tools

About 814,000,000 results (0.68 seconds)

[Very Nice and Inexpensive - Alfa Theater Hotel, Antwerp Traveller ...](#)
[www.tripadvisor.co.uk](#) > ... > [Antwerp hotels](#) > [Alfa Theater Hotel](#) ▾
★★★★★ Rating: 3.5 - 126 reviews - Price range: ££
Alfa Theater Hotel: **Very Nice and** Inexpensive - See 126 traveler reviews, 22 candid photos, and great deals for Antwerp, Belgium, at TripAdvisor.

[Very nice and clean hotel - Hotel Darival, Rome Traveller Reviews ...](#)
[www.tripadvisor.co.uk](#) > ... > [Lazio](#) > [Rome](#) > [Rome hotels](#) > [Hotel Darival](#) ▾
★★★★★ Rating: 4.5 - 27 reviews - Price range: ££
Hotel Darival: **Very nice and** clean hotel - See 27 traveler reviews, 11 candid photos, and great deals for Rome, Italy, at TripAdvisor.

[Very nice, centrally located and... - Hotel One Shot 23 - TripAdvisor](#)
[www.tripadvisor.co.uk](#) > ... > [Madrid](#) > [Madrid hotels](#) > [Hotel One Shot 23](#)
★★★★★ Rating: 4.5 - 78 reviews - Price range: ££
22 hours ago - The TV was **very nice and** big since the big mirror in front of my bed reflected the images. The room comes with shampoo, shower gel, cream, ...

[Very, very nice and modern hotel - The Pure, Frankfurt Traveller ...](#)
[www.tripadvisor.co.uk](#) > ... > [Frankfurt](#) > [Frankfurt hotels](#) > [The Pure](#) ▾
★★★★★ Rating: 4 - 360 reviews - Price range: ££
20 Sep 2013 - The Pure: Very, **very nice and** modern hotel - See 360 traveler reviews,

Lexica of emotion words/phrases (ctd) - from corpora

Google

"too expensive and"

Web Images Maps Shopping More ▾ Search tools

About 56,900,000 results (0.29 seconds)

Too expensive and not interesting - Barcelo Malaga ... - TripAdvisor
www.tripadvisor.co.uk > ... > Malaga > Malaga hotels > Barcelo Malaga ▾
★★★★★ Rating: 4.5 - 748 reviews - Price range: £
8 Oct 2013 - Barcelo Malaga: **Too expensive and** not interesting - See 748 traveler reviews, 322 candid photos, and great deals for Malaga, Spain, ...

The UK's new nuclear reactors: too expensive and not needed ...
www.greenpeace.org > Home > News > Blogs > Nuclear reaction ▾
The UK's new nuclear reactors: **too expensive and** not needed. Blogpost by Justin McKeating - October 21, 2013 at 14:17 10 comments. Looking closely at the ...

Are Edinburgh Fringe tickets too expensive? And gigs too cheap ...
www.thestage.co.uk/.../are-edinburgh-fringe-tickets-too-expensive-and-g... ▾
23 Aug 2013 - As the Edinburgh Fringe winds down this coming long weekend, the best reporter and critic on the events north of the border Lyn Gardner has ...

25/09/2013 Organic Licensing 'too expensive and too complex for ...
www.roythorne.co.uk/.../Organic-Licensing-'too-expensive-and-too-com... ▾
25 Sep 2013 - Organic licenses are '**too expensive and** too complex' for small producers says leading food law firm, Roythornes. Licenses in the UK are ...

Impact measurement 'hardly understood, too expensive and too ...

Two approaches to SA

- Lexicon-based
 - ◊ Binary
 - ◊ Gradable
- **Corpus-based/Supervised machine learning**

A corpus-based approach to SA

Idea: Mostly “supervised learning”: **corpora** of examples **annotated with sentiment** are used with machine learning algorithms to learn a classifier for each sentence/document. Corpora can be built:

- Manually: reliable, can be used as gold-standards
- From crowd-annotated resources, like Amazon Product Reviews (1-5 stars); Rotten Tomatoes, complaints.com, bitterlemons.com

Corpus: a collection of text segments (e.g. webpages, blog posts, reviews, tweets, etc) with humanly-annotated emotional indicators (e.g. positive, negative, etc).

E.g.: “If you are reading this because it is your darling fragrance, please wear it at home exclusively and tape the windows shut.” → {negative}

A corpus-based approach to SA - Corpora

Examples of corpora:

- Subjectivity corpus
 - ◇ 10,000 sentences: subjective/objective
 - ◇ Objective: IMDB plot summaries
 - ◇ Subjective: Rotten Tomatoes website.
- “Movie Review” corpus (Pang, Lee and Vaithyanathan, 2002):
 - ◇ 2,000 movie reviews (equal number of positive/negative)
 - ◇ Source: IMDB
- Many more:
<http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>

A corpus-based approach to SA - Features

Mostly words, but also other linguistic traits describing positive/negative examples:

- Words (unigrams)
- n-grams (sequences of n words)
- Emotions from words/phrases extracted from dictionaries
- Part-of-speech (POS) tags
- Syntactic patterns (e.g. sequences of POS tags)
- Language model scores: similarities to positive/negative corpora
- Negations

All automatically extracted from the corpus.

A corpus-based approach to SA - Machine Learning

Two steps:

- 1 Subjectivity classifier: first run binary classifier to identify and then eliminate objective segments
- 2 Sentiment classifier with remaining segments: learn how to combine and weight different attributes to make predictions. E.g. Naive Bayes

Pre-processing of corpus similar to IR:

- Remove HTML or other tags
- Remove stopwords
- Perform word stemming/lemmatisation
- etc.

Bing Liu and Lei Zhang (2012). A survey on opinion mining and sentiment analysis. Kluwer Academic Publishers:

http://www.cs.uic.edu/~lzhang3/paper/opinion_survey.pdf