UBFC UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ

ELLIADD
Édition · Littératures
Langage · Informatique
Arts · Didactique · Discours

Grey Sentiment Analysis

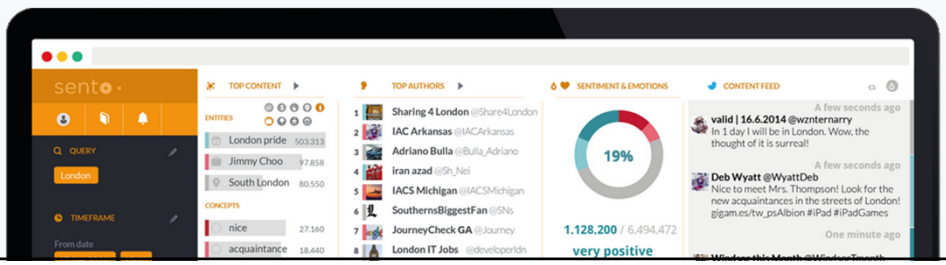
Liviu-Adrian Cotfas, Camelia Delcea, Ioan Roxin

liviu-adrian.cotfas@univ-fcomte.fr
liviu.cotfas@ase.ro

Sentiment Analysis

Sentiment analysis is a growing area of Natural Language Processing, commonly used to get insights from customer reviews, blogs and more recently from social media messages.

It requires a multidisciplinary approach, that combines elements from fields such as linguistics, psychology and artificial intelligence. It is used to determine whether a text expresses a **positive**, **negative** or **neutral perception**, also known as polarity.



"Everyone's customer will sooner or later express what they want on the Net. [...]. You just put your ear to the ground and listen. You probably want to ask questions too but that will be to get details, to fine tune — not to understand the picture, only to understand what particular shade of green the customer is seeing out of 3,500 shades of green."

Source: Sony [http://breakthroughanalysis.com/]

How?

Main approaches:

- Lexicon based;
- Machine learning based;
- Hybrid approaches.

Many Sentiment Analysis Lexicons…

What if?

we could combine existing lexicons to

Validate using a base-line sentiment analysis

Grey Sentiment Lexicons

MaxDiff-Twitter-Lexicon

+

VaderSentiment Lexicon

+

Sentiment140 Lexicon

⟶ Grey Sentiment Lexicons

Maxdiff-Twitter-Lexicon_-1to1

- number of tokens: 1,515
- contains: English words, emoticons, sentiment-related acronyms and initialisms (e.g. LOL), commonly used slang (e.g. nah)
- intensity range:
  –1 (extremely negative) to 1 (extremely positive)
- ex: the word "okay" has a positive valence of 0.376, "good" of 0.656, and "great" of 0.734
- url: http://www.purl.com/net/lexicons

VaderSentiment Lexicon

- number of tokens: 7,500
- contains: English words, emoticons, sentiment-related acronyms and initialisms (e.g. LOL), commonly used slang (e.g. nah)
- intensity range:
  –4 (extremely negative) to 4 (extremely positive)
- ex: the word "okay" has a positive valence of 0.9, "good" of 1.9, and "great" of 3.1
- url: https://github.com/cjhutto/vaderSentiment

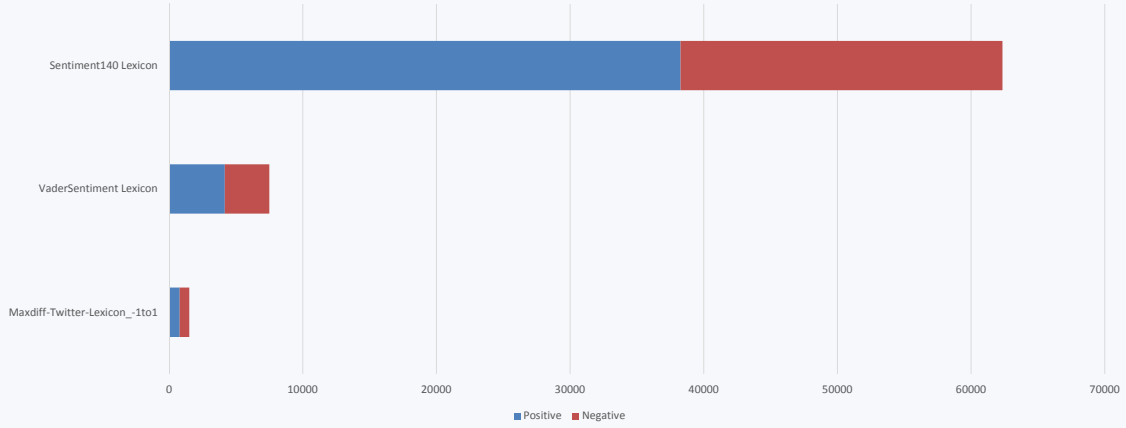
Sentiment140 Lexicon

- number of tokens: 62,468
- contains: English words, sentiment-related acronyms and initialisms (e.g. LOL), commonly used slang (e.g. nah)
- intensity range:
  –1 (extremely negative) to 1 (extremely positive)
- ex: the word "okay" has a positive valence of 0.056, "good" of 0.165, and "great" of 0.2354
- url: http://www.purl.com/net/lexicons

Comparison between the three lexicons

- **Positive** vs **Negative** terms

	Maxdiff-Twitter-Lexicon -1to1		VaderSentiment Lexicon		Sentiment140 Lexicon	
Positive	789	52%	4,171	56%	38,260	61%
Negative	726	48%	3,332	44%	24,089	39%



Grey Sentiment Lexicon

Steps:

1. normalize the values in all lexicons in the interval [-1,1];
2. [min, max]

Example (using all three lexicons):

	Maxdiff-Twitter-Lexicon -1to1	VaderSentiment Lexicon	Sentiment140 Lexicon	Grey Sentiment Lexicon
ok	0.376	0.225	0.056	[0.056 - 0.376]
good	0.656	0.475	0.165	[0.165 - 0.656]
great	0.734	0.775	0.2354	[0.2354 - 0.775]

Grey Sentiment Analysis

*"I am very **upset** and **disappointed** by this iphone update **failed** backup. @iphone"*

[-0.766, -0.4] + [-0.688,-0.44] + [-0.688,-0.334] = **[-0.688,-0.334]**

Results*

	Vander	Maxdiff	Sentiment 140	Vander ∩ MaxDiff	Vander ∩ MaxDiff2 ∩ Sentiment 140
TruePositive	429	436	324	451	445
TrueNegative	317	341	166	430	432
FalsePositive	188	164	339	75	73
FalseNegative	39	32	144	17	23
Precision	0.70	**0.73**	0.49	**0.86**	0.86
Recall	0.92	**0.93**	0.69	**0.96**	0.95
Accuracy	0.77	**0.80**	0.50	**0.91**	0.90

*obtained on the Sanders dataset: http://www.sananalytics.com/lab/twitter-sentiment/

Conclusions

- visible improvements in **precision**, **recall** and **accuracy** , when using Grey Lexicons;
- extremely noisy lexicons can negatively affect the performance of the resulting Grey Lexicon.

Further Research Directions

- integrate additional lexicons such as **SentiWordNet** and **SenticNet**;
- validate the approach using a state-of-the-art sentiment analysis algorithm;
- extend the proposed approach for **emotion analysis.**

Source code

https://github.com/lcotfas/GreySentimentAnalysis

Thank you!