

Entity Sentiment Extraction Using Text Ranking

John O'Neil

Attivio, Inc.

15 August 2012

An Example

I already hated AT&T. It's my fixed telephony and internet provider (because it has something of a monopoly on such services). I go through periods where my internet becomes intermittent, which AT&T refuses to acknowledge. . . I love love love my iPhone. It's my mini-computer on the go. I use it for texting, social sharing, photography, editing, keeping track of my calendar, storing contacts, finding directions, listening to music and podcasts, watching videos, reading, and blogging. Sometimes, I even make a phone call.

— <http://stumbledownunder.com/2012/01/07/using-my-beloved-iphone-in-australia/>

Entity Sentiment

- ▶ Entity extraction and document sentiment are well-known techniques.
- ▶ For many uses, it's important to assign sentiment to entities in a document, not to the document as a whole.
- ▶ How best to accomplish this?

TextRank (Mihalcea & Tarau 2004)

- ▶ Document as graph.
- ▶ Choose representation appropriately!
- ▶ Power iteration finds the dominant eigenvector.

Prerequisite: Entity Extraction

- ▶ A combination of statistical and rule-based approaches.
- ▶ We get the positions of the entity mentions in the document, and resolve matches.

Prerequisite: Document Sentiment

- ▶ Train on a corpus of positive and negative BOWs using your favorite linear classifier.
- ▶ This associates a (positive or negative) sentiment weight for each word (and optionally phrase) in the training corpus.

TextRank Highlights

- ▶ Document graph

 - Nodes** Words and entities

 - Edges** Between nearby words-entity pairs and word-word pairs.

 - Edge Weights** Word sentiment

- ▶ PageRank

- ▶ De-sparsified matrix

TextRank algorithm

Input: Initial set of vertex weights WS

Iterate until convergence:

$$WS(V_i) = (1 - d) + d * \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} WS(V_j)$$

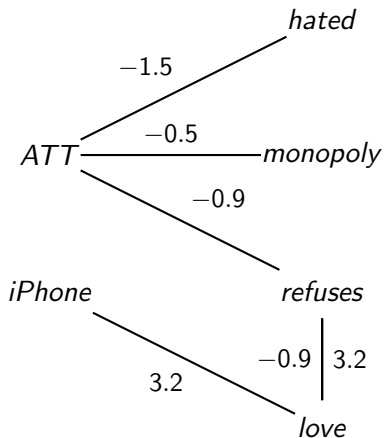
- ▶ w_{ij} is the weight of the edge going from vertex V_i to vertex V_j
- ▶ $In(V_i)$ are the edges that point to V_i
- ▶ $Out(V_i)$ are the edges that point away from V_i .
- ▶ d is a constant damping factor (typically 0.85).
- ▶ At convergence, WS contains the final sentiment weights.

An Example, Again

I already hated AT&T. It's my fixed telephony and internet provider (because it has something of a monopoly on such services). I go through periods where my internet becomes intermittent, which AT&T refuses to acknowledge. . . I love love love my iPhone. It's my mini-computer on the go. I use it for texting, social sharing, photography, editing, keeping track of my calendar, storing contacts, finding directions, listening to music and podcasts, watching videos, reading, and blogging. Sometimes, I even make a phone call.

— <http://stumbledownunder.com/2012/01/07/using-my-beloved-iphone-in-australia/>

Simple Text Graph



Simple Text Node Weights

	initial	final
AT&T	1.0	-1.25
iPhone	1.0	0.95
hated	1.0	-0.84
monopoly	1.0	-0.09
refuses	1.0	0.28
love	1.0	0.95

Main Uses of Entity Sentiment (for us)

Faceting Filling facets with entries relevant to the query.

Entities Creating metadata for entities, improving search.

Time Viewing entity sentiment changes over time.

Entity Sentiment Evaluation

- ▶ Without test corpus, compare systems:
 - TextRank** The one described here.
 - Baseline** System using document's sentiment for each entity in the document.
- ▶ Task: get most highly correlated (and anti-correlated) entity-&-sentiment pairs

Entity Sentiment Evaluation Corpus

- ▶ One day of the Moreover feed: 23 September 2011.
- ▶ Approximately 423,000 news articles in English, mostly U.S.

Top Headlines for 23 September, 2011

- ▶ *Idaho to seek waiver for No Child Left Behind law*
- ▶ *Spending Dispute Threatens U.S. Government Shutdown*
- ▶ *Faster than light? CERN findings bewilder scientists*
- ▶ *Saleh Returns to Yemen amid Increased Violence*
- ▶ *GOP Candidates Debate in Orlando; Audience Boos Gay Soldier*

Baseline: Top Document Co-occurrences on query *Obama*

	entity	log likelihood
Barack Obama	White House	4344.15
Barack Obama	Mitt Romney	3677.81
Barack Obama	Rick Perry	3612.59
Barack Obama	West Bank	3120.62
Barack Obama	Mahmoud Abbas	2879.53
Barack Obama	Jon Huntsman	2644.31
Barack Obama	Michele Bachmann	2526.38
Barack Obama	United States	2520.69
Barack Obama	Benjamin Netanyahu	2508.19
Barack Obama	Rick Santorum	2083.20

Baseline: Top Document Co-occurrences on query *Stephen Hill*

	entity	log likelihood
Stephen Hill	Rick Santorum	815.64
Stephen Hill	Gay Soldier	220.70
Stephen Hill	Rick Perry	195.20
Stephen Hill	Megyn Kelly	171.66
Stephen Hill	Ron Paul	165.22
Stephen Hill	Brian Williams	141.52
Stephen Hill	John Kerry	109.87
Stephen Hill	Mitt Romney	105.68
Stephen Hill	Herman Cain	90.21
Stephen Hill	Newt Gingrich	86.38

Baseline: Positive and Negative Entity Sentiment in Corpus

entity	%pos	%neg
Barack Obama	70.7	29.3
Congress	87.6	12.4
Michelle Bachmann	94.2	5.8
Rick Perry	79.7	20.3
Rick Santorum	82.5	17.5
Ron Paul	90.0	10.0
John Kerry	88.0	12.0
Mitt Romney	82.3	17.7
Herman Cain	85.1	14.9
Newt Gingrich	85.0	15.0
Ali Abdullah Saleh	38.9	61.1

TextRank: Positive and Negative Entity Sentiment in Corpus

entity	%pos	%neg
Barack Obama	46.25	53.75
Congress	46.0	54.0
Michelle Bachmann	0.0	100.0
Rick Perry	39.2	60.8
Rick Santorum	13.3	86.7
Ron Paul	34.5	65.5
Mitt Romney	70.5	29.5
Newt Gingrich	77.4	22.6
Ali Abdullah Saleh	5.6	94.4

TextRank: Top Same-Polarity Correlations on query *Obama*

	entity	log likelihood
Barack Obama	Idaho	314.57
Barack Obama	Eric Holder	134.66
Barack Obama	Arne Duncan	107.16
Barack Obama	Angela Merkel	103.03
Barack Obama	Education Department	74.15

TextRank: Top Opposite-Polarity Correlations on query *Obama*

	entity	log likelihood
Barack Obama	Mumbai Attackers	388.06
Barack Obama	Capitol Hill	282.10
Barack Obama	Republicans	144.94
Barack Obama	Congress	84.18
Barack Obama	Michele Bachmann	76.61

TextRank: Top Same-Polarity Correlations on query *Stephen Hill*

	entity	log likelihood
Stephen Hill	Gay Soldier	13.58

TextRank: Top Opposite-Polarity Correlations on query *Stephen Hill*

	entity	log likelihood
Stephen Hill	Fox News	114.39
Stephen Hill	Rick Santorum	116.08
Stephen Hill	Republican Debate	30.96
Stephen Hill	Rick Perry	18.43
Stephen Hill	Mitt Romney	3.98

TextRank: Top Same-Polarity Correlations on query *Congress*

	entity	log likelihood
Congress	Mitch Daniels	466.28
Congress	Senate	122.13
Congress	Democrats	117.57
Congress	Treasury Department	95.66
Congress	Sonia Gandhi	64.59

TextRank: Top Opposite-Polarity Correlations on query *Congress*

	entity	log likelihood
Congress	Capitol Hill	278.19
Congress	Americans	110.84
Congress	Barack Obama	54.55
Congress	Janet Napolitano	27.03
Congress	Senate	17.43

Conclusions & Future Directions

- ▶ Extraction of recognizably useful information.
- ▶ Need test corpus.

The End

Thanks! Questions?