A Machine Learning approach for Subjectivity Classification based on Positional and Discourse Features

Dr. David E. Losada

Centro Singular de Investigación en Tecnologías de la Información (CITIUS)
Universidad de Santiago de Compostela

david.losada@usc.es

Contents

1. Introduction
2. A Machine Learning approach for Subjectivity Classification based on Positional and Discourse Features
3. Method
4. Experiments
5. Related Work
6. Conclusions & Future Work

Contents

1. Introduction
2. A Machine Learning approach for Subjectivity Classification based on Positional and Discourse Features
3. Method
4. Experiments
5. Related Work
6. Conclusions & Future Work

Opinion Mining & Sentiment Analysis

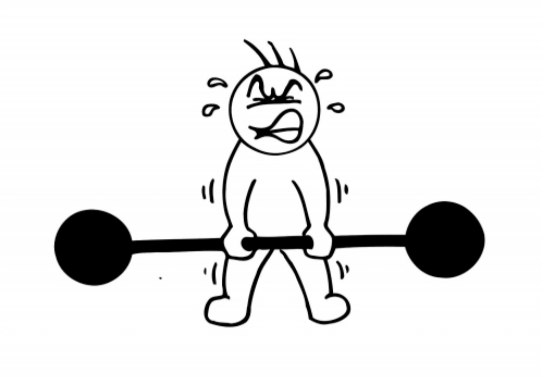
Opinion-rich resources

- Growing availability & popularity: online review sites, discussion forums, personal blogs, peer-to-peer networks, social networks, ...
- Opinions are very valuable: products/services, politics, ...
- But non-automated analyses (clipping services, field agents, adhoc research): can't keep pace.
- OM & SA technology: potentially wide industrial impact

But...

OM & SA technology

- Still not ready for prime time!
- Modest levels of effectiveness



Topic retrieval vs Opinion retrieval

Topic retrieval

- Estimating topicality is somehow easier
- Keyword-based approaches work reasonably well
- Effective retrieval algorithms
- Massive success:

Topic retrieval vs Opinion retrieval

Opinion retrieval

- Sentiment classification is harder
  - Search for on-topic opinions: difficult passage-level task
  - Locate key sentiments is challenging
  - Deal with irony, sarcasm, etc.
  - Context and Language dependent!
- Keyword-based approaches fail

Objective vs Subjective

Skype 2.0 eats its young

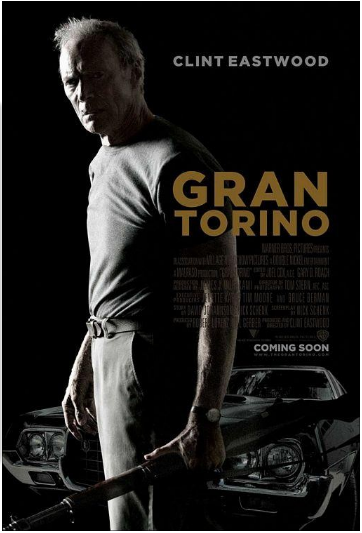
The elaborate press release and WSJ review while impressive don't help mask the fact that, Skype is short on new ground breaking ideas. Personalization via avatars and ring-tones · · · big new idea? Not really. Phil Wolff over on Skype Journal puts it nicely when he writes, "If you've been using Skype, the Beta version of Skype 2.0 for Windows won't give you a new Wow! experience."· · ·

Skype Launches Skype 2.0 Features Skype Video

Skype released the beta version of Skype 2.0, the newest version of its software that allows anyone with an Internet connection to make free Internet calls. The software is designed for greater ease of use, integrated video calling, and · · ·

Sentiment classification

Gran Torino also includes a few easy outs built into the story ... And even without those easy outs, the storytelling's fairly obvious ... Gran Torino is a curdled mess, politically ... but considering that Gran Torino's heading towards the sunset of Eastwood's acting career, that's a good enough reason to watch it go by.



Sentiment classification

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, . . . [they] act wacky as hell . . . the ninth floor of hell . . . a cheap [beep] movie . . . The plot is such a mess that it's terrible. But I loved it.

Contents

1 Introduction

2 A Machine Learning approach for Subjectivity Classification based on Positional and Discourse Features

3 Method

4 Experiments

5 Related Work

6 Conclusions & Future Work

IRFC 2013 paper

Jose M. Chenlo, David E. Losada. *A Machine Learning approach for Subjectivity Classification based on Positional and Discourse Features*, 6th Information Retrieval Facility Conference, Limassol, Cyprus, October 2013.

Sentiment classification

Global (doc-level) methods

- Ignore the sequence of opinions
- Rough doc-level estimations
- Poor effectiveness in searching for pos & neg docs

Main motivation

Inject more advanced evidence:

- Structural aspects of natural language text (discourse)
- Position
- Sentence-level estimation

Contents

1. Introduction
2. A Machine Learning approach for Subjectivity Classification based on Positional and Discourse Features
3. Method
4. Experiments
5. Related Work
6. Conclusions & Future Work

Detecting Subjectivity

Skype 2.0 eats its young

The elaborate press release and WSJ review while impressive don't help mask the fact that, Skype is short on new ground breaking ideas. Personalization via avatars and ring-tones · · · big new idea? Not really. Phil Wolff over on Skype Journal puts it nicely when he writes, "If you've been using Skype, the Beta version of Skype 2.0 for Windows won't give you a new Wow! experience."· · ·

Skype Launches Skype 2.0 Features Skype Video

Skype released the beta version of Skype 2.0, the newest version of its software that allows anyone with an Internet connection to make free Internet calls. The software is designed for greater ease of use, integrated video calling, and · · ·

Sentence Features I

- Document-level sentiment classication is too crude for most applications

- Sentence level ⇒ a more advanced analysis of sentiments
- Positional information and discourse structure:
  - Key sentiments: specific locations
  - Rhetorical roles of text segments can effectively guide the opinion detection process
  - subjectivity of a document being not so much conveyed by the sentiment-carrying words that people use, but rather by the way in which these words are used

Sentence Features II

- Unigram & Bigrams
- Sentiment Lexicon
- Rhetorical Structure Theory
- Length
- Position

Unigram/Bigrams & Sentiment Lexicon Features

Unigram/Bigrams

Binary features based on the appearance of unigrams and bigrams in the sentence.

Sentiment Lexicon (OpinionFinder)

Sentiment-bearing terms that occur in the sentence.

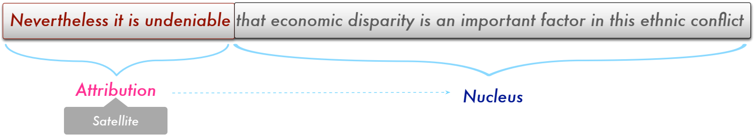
- Number and percentage of opinionated terms in the text.
- Number and percentage of interrogations and exclamations.

Rhetorical Features I

- Subjectivity estimation using (sentence-level) discourse structure
- Rhetorical Structure Theory (RST):
  - Sentences split into nucleus+ satellite

Nevertheless it is undeniable that economic disparity is an important factor in this ethnic conflict

- Different rhetorical relations: attribution, background, cause, contrast, elaboration, ...



Rhetorical Features II

Relation	Description
attribution	Clauses containing reporting verbs or cognitive predicates related to reported messages presented in nuclei.
background	Information helping a reader to sufficiently comprehend matters presented in nuclei.
cause	An event leading to a result presented in nuclei.
comparison	Clauses presenting matters which are examined along with matters presented in nuclei in order to establish similarities and dissimilarities.
condition	Hypothetical, future, or otherwise unrealized situations, the realization of which influences the realization of nucleus matters.
contrast	Situations juxtaposed to situations in nuclei, where juxtaposed situations are considered as the same in many respects, yet differing in a few respects, and compared with respect to one or more differences.
elaboration	Rhetorical elements containing additional detail about matters presented in nuclei.
enablement	Rhetorical elements containing information increasing a readers' potential ability of performing actions presented in nuclei.
evaluation	An evaluative comment about the situation presented in the associated nucleus.
explanation	Justifications or reasons for situations presented in nuclei.
joint	No specific relation is assumed to hold with the matters presented in the associated nucleus.
temporal	Clauses describing events with a specific ordering in time with respect to events described in nuclei.

Rhetorical Features III

Contrast relationships

- Contrast of the statements presented in the satellite and nucleus
- Evidence in favour of subjectivity?

Contrast

A degree of selfishness in capitalist countries seems to be part of the ideology, but one of the great lessons of this bloody 20th century was that pure self-interest needs to be tempered by a contribution to the more general good

Temporal relationships

- Evidence in favour of objectivity?

Temporal

Pakistan detonated a series of nuclear devices last month after India surprised the world with its tests

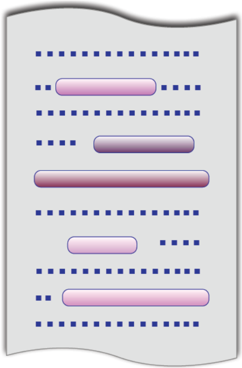
Length Features

- Length of the sentence
- Length of the nucleus
- Length of the satellite

Positional Features

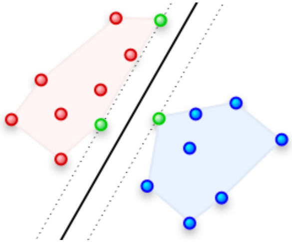
- Positional features could be highly indicative of opinions
  - Opinions at the end?

- Absolute position of the sentence within the document
- Relative position of the sentence within the document
- Number of sentences in the document



Classification: Support Vector Machines (SVMs)

- A two-class (subjective vs. non-subjective) classification problem
- Highly effective in many learning problems
- Linear classifiers: facilitates the analysis
  - Weights of the separating hyperplane can be used to assess the relevance of each feature



Contents

1 Introduction

2 A Machine Learning approach for Subjectivity Classification based on Positional and Discourse Features

3 Method

4 Experiments

5 Related Work

6 Conclusions & Future Work

Benchmark

- Collection of news: NTCIR-7 English MOAT Research collection
- Annotated data at sentence level (relevance and subjectivity)
- The labels were produced by three different assessors
  - Majority rule
- 3584 sentences: 2697 judged as objective and 887 judged as subjective
  - 2218 unigrams and 2812 bigrams

baseline

- OpinionFinder
  State-of-the-art sentence level subjectivity classifier

Results

- Most of our methods outperform OF
- Our method with all features performs the best
- Positional features seem to be important
- Sentiment lexicon and length also contribute to improve the basic classifiers
  - Precision vs. Recall

	Precision	Recall	F1
OpinionFinder	.4420	.4126	.4268
unigrams	.4926	.3855	.4325
+ Rhetorical	.4903	.4140	.4489
+ Positional	.4716	**.5033**	**.4869**
+ Length	.4571	.4846	.4704
+ Sent. Lex.	**.5077**	.4513	.4778
+ All	.4892	.4822	.4857
unigrams & bigrams	.5410	.3591	.4317
+ Rhetorical	.4903	.3576	.4248
+ Positional	.5045	.4573	.4797
+ Length	.4806	.4464	.4629
+ Sent. Lex.	**.5517**	.3883	.4558
+ All	.4858	**.5150**	**.5000**

Results

- Rhetorical information alone modestly improves performance
- Good in combination with other features (e.g., opinion lexicon features)
  - RST can modulate the influence of lexicon-based information
- Some relations are highly indicative of subjectivity

	Precision	Recall	F1
OpinionFinder	.4420	.4126	.4268
unigrams	.4926	.3855	.4325
+ Rhetorical	.4903	.4140	.4489
+ Positional	.4716	**.5033**	**.4869**
+ Length	.4571	.4846	.4704
+ Sent. Lex.	**.5077**	.4513	.4778
+ All	.4892	.4822	.4857
unigrams & bigrams	.5410	.3591	.4317
+ Rhetorical	.4903	.3576	.4248
+ Positional	.5045	.4573	.4797
+ Length	.4806	.4464	.4629
+ Sent. Lex.	**.5517**	.3883	.4558
+ All	.4858	**.5150**	**.5000**

Most Discriminative Features

- The two most discriminative features are the number of negative words and the position of the sentence in the document
- The most discriminative vocabulary features are the unigrams *objections* and *expressed*
- Personal pronouns (e.g., they, I) highly discriminative
- Interrogation/exclamations is indicative of objectivity in this dataset

rank	w_i	feature	feature set
1	3.0439	#Neg	Opinion
2	2.4448	nSent	Position
3	-2.4210	#ExcInt	Opinion
4	2.3093	objections	vocab.
5	2.2380	expressed	vocab.
6	2.2355	they are	vocab.
7	-2.2031	nSentsDoc	Length
8	2.1838	globalisation	vocab.
9	2.1239	actions	vocab.
10	2.0839	Nor	vocab.
11	2.0037	notably	vocab.
12	-1.9996	weather	vocab.
13	1.9034	means	vocab.
14	1.8829	something	vocab.
15	1.8137	I	vocab.

rank	w_i	feature	feature set
16	-1.8026	market	vocab.
17	-1.7575	expected	vocab.
18	-1.7527	key	vocab.
19	-1.7205	will have	vocab.
20	1.7190	America	vocab.
21	1.7002	#PosNorm	Opinion
22	1.6894	should	vocab.
23	1.6823	investors	vocab.
24	-1.6593	financial	vocab.
25	-1.6522	world economy	vocab.
26	-1.6449	to use	vocab.
27	1.6324	said in	vocab.
28	1.6182	programs	vocab.
29	1.6095	ministers	vocab.
30	1.6087	US economy	vocab.

Most Discriminative Non-vocabulary Features

- The most discriminative features tend to be terms provided by OF lexicon

rank	w_i	feat	
1	3.0439	#Neg	Op.
2	2.4448	nSent	Pos.
3	-2.4210	#ExclInt	Op.
4	-2.2031	nSentsDoc	Leng.
5	1.7002	#PosNorm	Op.
6	1.5764	#Pos	Op.
7	1.4859	#NegNorm	Op.
8	-1.4224	#ExclIntNorm	Op.
9	1.3025	has *Evaluation* sat.	RST
10	-1.2566	nSentNorm	Pos.
11	0.9867	has *Attribution* sat.	RST
12	-0.8718	has *Temporal* sat.	RST
13	-0.8442	has *Background* sat.	RST

rank	w_i	feat	
14	0.4591	has *Comparison* sat.	RST
15	0.4220	lengthSat	Leng.
16	-0.3927	has *Manner* sat.	RST
17	-0.3338	has *Cause* sat.	RST
18	-0.3034	lengthNuc	Leng.
19	-0.2612	has *Contrast* sat.	RST
20	0.2319	has *Condition* sat.	RST
21	-0.1997	has *Enablement* sat.	RST
22	0.1643	lengthSent	Leng.
23	-0.1635	has *Explanation* sat.	RST
24	-0.1170	has *Elaboration* sat.	RST
25	0.1112	has *Joint* sat.	RST
26	-0.0924	hasSat	RST

Most Discriminative Non-vocabulary Features

- *evaluation*, *attribution* and *comparison* => subjectivity
  - e.g., *attribution* statements when the author of the article writes about others' opinions

According to the new CEO, *the future of the company is brilliant*

rank	w_i	feat	
1	3.0439	#Neg	Op.
2	2.4448	nSent	Pos.
3	-2.4210	#ExclInt	Op.
4	-2.2031	nSentsDoc	Leng.
5	1.7002	#PosNorm	Op.
6	1.5764	#Pos	Op.
7	1.4859	#NegNorm	Op.
8	-1.4224	#ExclIntNorm	Op.
9	1.3025	has *Evaluation* sat.	RST
10	-1.2566	nSentNorm	Pos.
11	0.9867	has *Attribution* sat.	RST
12	-0.8718	has *Temporal* sat.	RST
13	-0.8442	has *Background* sat.	RST

rank	w_i	feat	
14	0.4591	has *Comparison* sat.	RST
15	0.4220	lengthSat	Leng.
16	-0.3927	has *Manner* sat.	RST
17	-0.3338	has *Cause* sat.	RST
18	-0.3034	lengthNuc	Leng.
19	-0.2612	has *Contrast* sat.	RST
20	0.2319	has *Condition* sat.	RST
21	-0.1997	has *Enablement* sat.	RST
22	0.1643	lengthSent	Leng.
23	-0.1635	has *Explanation* sat.	RST
24	-0.1170	has *Elaboration* sat.	RST
25	0.1112	has *Joint* sat.	RST
26	-0.0924	hasSat	RST

Most Discriminative Non-vocabulary Features

- *temporal* and *background* => objectivity
  - *temporal* statements tend to be objective and are often used to locate events in time

The day after the attacks, we saw immediate cancellations

  - *background* statements indicates the nature of the information presented in nucleus

Culturally they are divided into peranakan and totok

rank	w_i	feat	
1	3.0439	#Neg	Op.
2	2.4448	nSent	Pos.
3	-2.4210	#ExclInt	Op.
4	-2.2031	nSentsDoc	Leng.
5	1.7002	#PosNorm	Op.
6	1.5764	#Pos	Op.
7	1.4859	#NegNorm	Op.
8	-1.4224	#ExclIntNorm	Op.
9	1.3025	has *Evaluation* sat.	RST
10	-1.2566	nSentNorm	Pos.
11	0.9867	has *Attribution* sat.	RST
12	-0.8718	has *Temporal* sat.	RST
13	-0.8442	has *Background* sat.	RST

rank	w_i	feat	
14	0.4591	has *Comparison* sat.	RST
15	0.4220	lengthSat	Leng.
16	-0.3927	has *Manner* sat.	RST
17	-0.3338	has *Cause* sat.	RST
18	-0.3034	lengthNuc	Leng.
19	-0.2612	has *Contrast* sat.	RST
20	0.2319	has *Condition* sat.	RST
21	-0.1997	has *Enablement* sat.	RST
22	0.1643	lengthSent	Leng.
23	-0.1635	has *Explanation* sat.	RST
24	-0.1170	has *Elaboration* sat.	RST
25	0.1112	has *Joint* sat.	RST
26	-0.0924	hasSat	RST

Contents

1. Introduction
2. A Machine Learning approach for Subjectivity Classification based on Positional and Discourse Features
3. Method
4. Experiments
5. Related Work
6. Conclusions & Future Work

RST for polarity detection

- What about RST for polarity estimation?

Although it was great to see Brad Pitt fall off a cliff, this movie was terrible

- :) or :(

- Polarity estimation using (sentence-level) discourse structure
- e.g., contrast Relationship => shift the score of the satellite

- Preliminary results published in collaboration with the Erasmus University of Rotterdam (Alexander Hogenboom)
  - *Jose M Chenlo, Alexander Hogenboom and David E. Losada, Sentiment-based Ranking of Blog Posts using Rhetorical Structure Theory*, NLDB 2013, Manchester (UK)

Contents

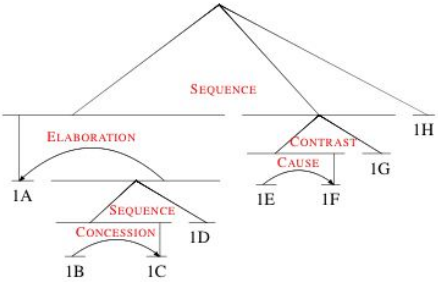
1. Introduction
2. A Machine Learning approach for Subjectivity Classification based on Positional and Discourse Features
3. Method
4. Experiments
5. Related Work
6. Conclusions & Future Work

Conclusions

- We explored the importance of sentence features in fine-grained subjectivity classification processes
  - e.g., positional or rhetorical features
- These features are valuable and can be combined with more classical methods based on unigrams, bigrams and subjectivity lexicon

Future Work

- Validate these findings against other datasets
- Study more advanced ways to combine features and classifiers
- Inter-sentence RST analysis



[Yesterday, the delegates chose their new representative.]^{1A} [Even though Smith received only 24 votes,]^{1B} [he accepted the election with a short speech.]^{1C} [Then the assembly applauded for three minutes.]^{1D} [Due to the upcoming caucus meeting,]^{1E} [the subsequent discussion was very short.]^{1F} [Nonetheless the most pressing questions could be resolved.]^{1G} [The meeting was closed at 7pm.]^{1H}