Combining Strengths, Emotions and Polarities for Boosting Twitter Sentiment Analysis

Felipe Bravo-Márquez[1], Marcelo Mendoza[2,3], Bárbara Poblete[1,3]

[1]Universidad de Chile
[2]Universidad Técnica Federico Santa María
[3]Yahoo! Research Latin America

KDD

WISDOM
Chicago, August 12th, 2013

Contribution

What?

We boost sentiment classification performance.

How?

Combining polarity, emotion, and strength oriented sentiment analysis lexical resources with existing opinion mining methods.

Why?

To improve two major sentiment analysis tasks: 1) Subjectivity classification, and 2) Polarity classification.

A taxonomy of lexical resources for sentiment analysis

1. **Polarity**: Polarity-oriented lexical resources are composed by lists of positive and negative words.
2. **Emotion**: Emotion-oriented lexical resources should provide a list of words or expressions marked according to different emotion states.
3. **Strength**: Strength-oriented lexical resources provide lists of opinion words together with intensity scores regarding an opinion dimension.

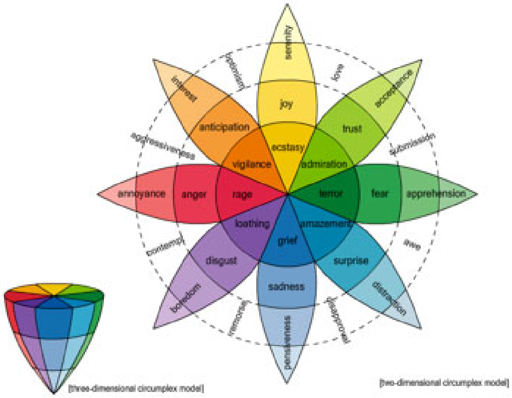
Strength, emotion and polarity-based resources are complementary

These are different dimensions of the same problem!

A taxonomy of lexical resources for sentiment analysis

1. **OpinionFinder Lexicon (OPF)**: polarity-oriented, list of sentences and single words with polarity tags. Size: 6,884 English words.
2. **AFINN Lexicon**: strength-oriented, list of positive words scored from 1 to 5 and negative words from -1 to -5. Size: 2,477 English words.
3. **SentiWordNet 3.0 (SWN)**: strength-oriented, list of synsets scored in $[-1, 1]$, Size: 147,306 wordnet synsets tagged.
4. **NCR Lexicon**: emotion-oriented, list of words with emotional tags according to the Plutchik's wheel of emotions. Size: 14,182 English words.

Plutchik's wheel of emotions



Lexical Resource Analysis

	SWN3	NCR	AFINN	OPFIND
SWN3	147,306			
NCR	13,634	14,182		
AFINN	1,783	1,207	2,476	
OPFIND	6,199	3,596	1,245	6,884
Total Distinct Words	149,114			

Table: Intersection of words between different Lexical Resources

Lexical Resource Analysis

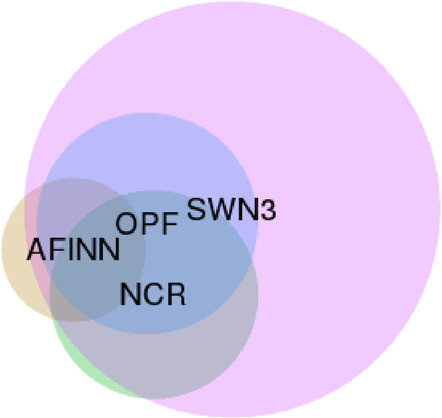


Figure: Non-neutral words interaction Venn diagram

Lexical Resource Analysis

	SWN3	NCR	AFINN	OPFIND
SWN3	33,313			
NCR	2,932	3,071		
AFINN	1,203	721	1,871	
OPFIND	3,703	1,658	900	4,311
Total Distinct Words	34,649			

Table: Intersection of non-neutral words

Lexical Resource Analysis

word	SWN3	AFINN	OPFIND	NCR
abuse	-0.51	-3	negative	ang,disg,fear,sadn
adore	0.38	3	positive	ant, joy, trust
cheer	0.13	2	positive	ant, joy, surp, trust
shame	-0.52	-2	negative	digs,fear,sadn
stunned	-0.31	-2	positive	fear, surpr
sympathy	-0.13	2	negative	sadn
trust	0.23	1	positive	trust
ugly	-0.63	-3	negative	disg
wonderful	0.75	4	positive	joy, surp, trust

Table: Sentiment Values of Words included in all the Resources

Combining lexicons

How?

We propose a supervised approach for which we model tweets as vectors of sentiment features.

Machine learning favors:

1. To explore the descriptive/discriminative power of each feature.
2. To test if a combination gets better results than each isolated resource.

Features

Feature	Description	Range
SSPOL	SentiStrength label (negative, neutral, positive)	$\{-1, 0, +1\}$
S140	S140 label (negative, neutral, positive)	$\{-1, 0, +1\}$
OPW	number of positive words that matches OpinionFinder	$\{0, 1, ..., n\}$
ONW	number of negative words that matches OpinionFinder	$\{0, 1, ..., n\}$
SSP	SentiStrength score for the positive category	$\{1, ..., 5\}$
SSN	SentiStrength score for the negative category	$\{-5, ..., -1\}$
SWP	$\sum$ of the scores for the positive words that matches SW3	$\{0, \ldots, n\}$
SWN	$\sum$ of the scores for the negative words that matches SW3	$\{0, ..., n\}$
APO	$\sum$ of the scores for the positive words that matches AFINN	$\{0, ..., n\}$
ANE	$\sum$ of the scores for the negative words that matches AFINN	$\{-n, ..., 0\}$
NJO	number of words that matches the joy word list of NCR	$\{0, 1, ..., n\}$
NTR	... matches the trust word list of NCR	$\{0, 1, ..., n\}$
NSA	... matches the sadness word list of NCR	$\{0, 1, ..., n\}$
NANG	... matches the anger word list of NCR	$\{0, 1, ..., n\}$
NSU	... matches the surprise word list of NCR	$\{0, 1, ..., n\}$
NFE	... matches the fear word list of NCR	$\{0, 1, ..., n\}$
NANT	... matches the anticipation word list of NCR	$\{0, 1, ..., n\}$
NDIS	... matches the disgust word list of NCR	$\{0, 1, ..., n\}$

Experimental design

Tasks and datasets

Tasks: Subjectivity (Neu), Polarity (Pol).
Datasets: Stanford Twitter Sentiment (STS), Sanders.

	STS	Sanders
#negative	177	636
#neutral	139	2,429
#positive	182	560
#total	498	3,625

Datasets

Subjectivity	STS	Sanders
#neutral	139	1,196
#subjective	139	1,196
#total	278	2,392

Polarity	STS	Sanders
#negative	177	560
#positive	177	560
#total	354	1,120

Balanced datasets

Information gain

Scope	Feature	Subjectivity STS	Subjectivity Sanders	Polarity STS	Polarity Sanders
Polarity	SSPOL	**0.179**	0.089	**0.283**	**0.192**
	S140	**0.103**	0.063	**0.283**	**0.198**
	OPW	0.088	0.024	0.079	0.026
	ONW	0.097	0.024	**0.135**	0.075
Strength	SSP	0.071	0.037	**0.200**	**0.125**
	SSN	0.090	0.044	**0.204**	**0.118**
	SWN	0.090	0.023	**0.147**	0.089
	SWP	**0.104**	0.030	0.083	0.015
	APO	0.088	0.024	0.079	0.026
	ANE	**0.134**	0.048	**0.200**	**0.143**
Emotion	NJO	0.000	0.000	0.055	0.065
	NTR	0.000	0.000	0.000	0.000
	NSA	0.000	0.017	0.000	0.056
	NANG	0.000	0.016	0.046	0.055
	NSU	0.000	0.000	0.000	0.017
	NFE	0.000	0.008	0.039	0.024
	NANT	0.000	0.000	0.000	0.000
	NDIS	0.000	0.014	0.056	0.030

Feature selection (CFS)

	Neu.STS	Neu.San	Pol.STS	Pol.San
ANE	✓	✓	✓	✓
APO	✓		✓	✓
ONW	✓		✓	✓
OPW	✓			
NJO				✓
S140	✓	✓	✓	✓
SSN			✓	✓
SSP			✓	
SSPOL	✓	✓	✓	✓
SWN	✓		✓	✓
SWP	✓	✓		

Table: Selected Features by CFS algorithm

Subjectivity performance (10 fold CV)

		STS		Sanders	
Features	Methods	accuracy	F_1	accuracy	F_1
Baseline.1	Sent140	0.655	0.538	0.615	0.524
Baseline.2	SSPOL	0.734	0.747	0.659	0.690
All	CART	0.694	0.693	0.686	0.685
	Naive Bayes	0.737	0.714	0.649	0.583
	SVM	0.763	0.761	**0.701**	0.705
FS	CART	0.730	0.727	0.677	**0.717**
	Naive Bayes	0.759	0.733	0.651	0.581
	SVM	0.773	**0.780**	0.680	0.696
Polarity	CART	0.734	0.747	0.677	**0.717**
	Naive Bayes	0.748	0.737	0.671	0.655
	SVM	0.759	0.756	0.674	0.713
Strength	CART	0.719	0.713	0.661	0.669
	Naive Bayes	0.766	0.741	0.636	0.558
	SVM	**0.777**	0.760	0.694	0.703
Emotion	CART	0.579	0.471	0.586	0.490
	Naive Bayes	0.579	0.480	0.573	0.428
	SVM	0.597	0.552	0.594	0.532

Polarity performance (10 fold CV)

		STS		Sanders	
Features	Methods	accuracy	F_1	accuracy	F_1
Baseline.1	SentiStrength	0.777	0.781	0.733	0.732
Baseline.2	AFINN	0.771	0.758	0.713	0.691
All	CART	0.788	0.788	0.780	0.789
	Naive Bayes	0.794	0.807	0.774	0.794
	SVM	0.808	0.808	**0.801**	**0.810**
Best.First	CART	0.791	0.797	0.789	0.789
	Naive Bayes	0.811	0.822	0.788	0.802
	SVM	**0.816**	**0.823**	0.792	0.804
Polarity	CART	0.802	0.804	0.779	0.797
	Naive Bayes	0.805	0.811	0.756	0.766
	SVM	0.799	0.810	0.776	0.797
Strength	CART	0.780	0.778	0.705	0.720
	Naive Bayes	0.780	0.794	0.762	0.787
	SVM	0.799	0.798	0.779	0.793
Emotion	CART	0.684	0.729	0.658	0.691
	Naive Bayes	0.641	0.704	0.654	0.720
	SVM	0.624	0.668	0.656	0.695

Remarks

1. Opinions are multidimensional objects.
2. Polarity tagging is a kind of projection of a multidimensional object into a two dimensional simplex.
3. Lexicon combination favors the reduction of information loss.