Exploring Sentiment in Social Media: Bootstrapping Subjectivity Clues from Multilingual Twitter Streams

Svitlana Volkova, Theresa Wilson, David Yarowsky

Center for Language and Speech Processing, Johns Hopkins University, Human Language Technology Center of Excellence

Motivation

The language people use to express opinions and sentiment in social media is extremely diverse:

- informal, with abbreviations and misspellings;
- funny, creative, and entertaining;
- topics change rapidly;
- people to invent new words.

– Statistical models trained on social media data may degrade over time ("occupy" has not been indicative of sentiment before 2011).

– General, domain-independent sentiment lexicons exist for limited languages and have low coverage.

Bootstrapping has been used for learning sentiment lexicons in other domains (Turney and Littman, 2002; Banea et al., 2008) but previous works mainly relied on:

- existing NLP tools, e.g., syntactic parsers (Wiebe, 2000);
- information extraction (IE) tools (Riloff and Wiebe, 2003);
- rich lexical resources – WordNet (Esuli and Sebastiani, 2006).

Our approach:

I. handles creativity and dynamic nature of social media (SM);
II. does not rely on any language-dependent tools;
III. scales to many under-explored languages and dialects in SM;
IV. allows real-time sentiment analysis in a streaming mode.

Data and code available at:
http://www.cs.jhu.edu/~svitlana/

Approach

Goal: create Twitter-specific lexicons via bootstrapping sentiment-bearing terms from multilingual Twitter streams.

Assumptions:

- sentiment-bearing terms of similar orientation co-occur (opposite do not co-occur) at the tweet level;
- the co-occurrence of domain-specific and domain-independent terms is a signal of subjectivity.

English Initial Lexicon: strongly subjective terms from MPQA lexicon.

Russian, Spanish Initial Lexicons:

1) translated terms from MPQA lexicon,
2) expanded verbs, adverbs and adjectives using morphological rules,
3) propagated term polarity and verified it using annotator judgments.

Iterative bootstrapping procedure with parameters:

- Θ_{pr} probability of a new term appearing with strongly subjective terms, Θ_k terms added per iteration and Θ_{freq} term frequency.

Subjectivity:

$$p^{subj}(w) = \frac{c\left(w, L_B(\bar{\theta})\right)}{c(w)}$$

Polarity:

$$p^{pos}(w) = \frac{c\left(w, L_B^{pos}(\bar{\theta})\right)}{c(w)}$$

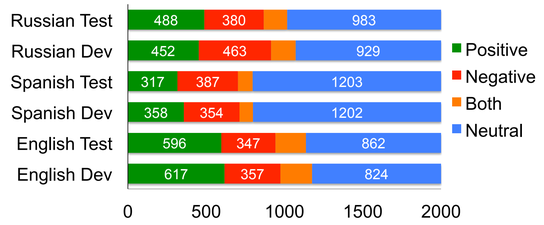
$$p^{pos}\left(w \middle| L_B(\bar{\theta})\right) + p^{neg}\left(w \middle| L_B(\bar{\theta})\right) = 1$$

- Optimize Θ parameters using grid search on DEV data by maximizing F-measure for subjectivity classification. English Θ = [0.7, 5, 50], Spanish and Russian Θ = [0.65, 3, 50].

Experiments

Data:

- 1M unlabeled tweets to bootstrap.
- 2K labeled tweets (DEV) used for parameter tuning.
- 2K labeled tweets (TEST) used to evaluate the quality of the lexicons.



Lexicon Evaluations: comparing with existing dictionary-expanded lexicons using rule-based subjectivity and polarity classifiers (Riloff, 2003).

I. English: 8K strongly subjective terms from SentiWordNet 3.0 (Esuli and Sebastiani, 2006).

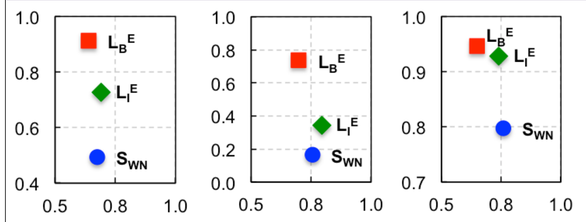
II. Spanish: 1.5K full and 2.5K medium strength terms from sentiment lexicons constructed by Perez-Rosas et al. (2012).

III. Russian: 5K terms from the lexicon constructed by Chetviorkin and Loukachevitchet (2012).

Error Analysis:

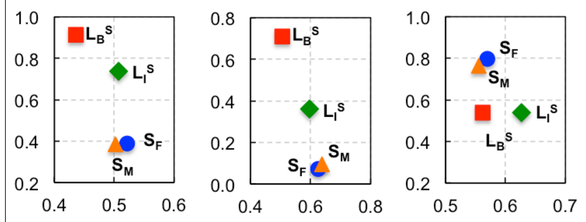
- POS tagging reduces FN errors for subjectivity classification (adjective, adverbs and verbs).
- FP errors for subjectivity classification: some terms are weakly subjective, can be used as subjective and neutral.
- Polarity classification errors: relying on positive or negative polarity but some term have mixed polarity.

Results



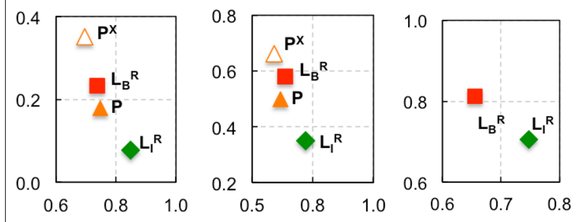
Precision (x-axis), recall (y-axis) for English: L_I^E = initial lexicon (5.1K), L_B^E = bootstrapped lexicon (21.5K), S_{WN} = SentiWordNet.

	$F_{subj} \geq 1$	$F_{subj} \geq 2$	F_{pol}
S_{WN}	0.57	0.27	0.78
L_I^E	0.71	0.48	0.82
L_B^E	0.75	0.72	0.78



Precision and recall for Spanish: L_I^S = initial (8.1K), L_B^S = bootstrapped lexicon (22.3K), S_F= full and S_M= medium strength lexicons.

	$F_{subj} \geq 1$	$F_{subj} \geq 2$	F_{pol}
S_F	0.44	0.17	0.64
S_M	0.47	0.13	0.66
L_I^S	0.59	0.45	0.58
L_B^S	0.59	0.59	0.55



Precision and recall for Russian: L_I^R = initial (3.7K), L_B^R = bootstrapped lexicon (10.8K), P = external and P_X = expanded lexicons.

	$F_{subj} \geq 1$	$F_{subj} \geq 2$	F_{pol}
P	0.55	0.29	-
P_X	0.62	0.47	-
L_I^R	0.46	0.13	0.73
L_B^R	0.61	0.35	0.73