NRC-Canada-2014: Recent Improvements in the Sentiment Analysis of Tweets

Xiaodan Zhu, Svetlana Kiritchenko, and Saif M. Mohammad

In Proceedings of the eighth international workshop on Semantic Evaluation Exercises (SemEval-2014), August, 2014, Dublin, Ireland

NRC·CNRC

Information and Communication Technologies

Sentiment Analysis of Term in Context: Task A

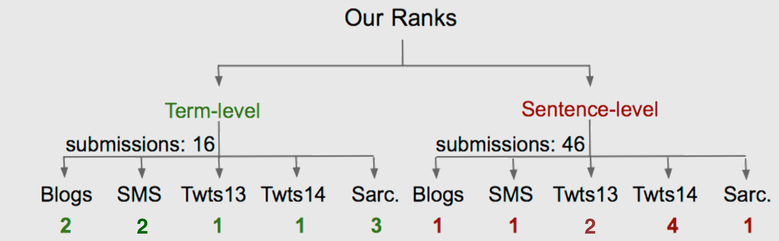
Polarity of the target: positive, negative, neutral?

Tweet: The movie is visually spectacular.
target is positive

Tweet: The movie felt like a slow documentary.
target is negative

Tweet: The NatGeo documentary was fascinating.
target is neutral

We present a sentiment analysis system which determines both term- and message-level polarity of a message.



Sentiment Analysis of Message: Task B

Polarity of the message: positive, negative, neutral?

Tweet: The movie is visually spectacular.
target is positive

Tweet: The movie felt like a slow documentary.
target is negative

Tweet: The NatGeo documentary is at 7pm.
target is neutral

The NRC-Canada-2014 System

List of Features

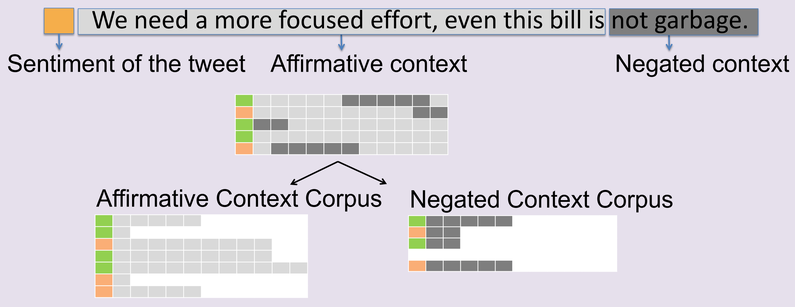
sentiment lexicon	# sentiment words, score
word n-grams	spectacular, like documentary
char n-grams	spect, docu, visua
part of speech	N: 5, V: 2, A:1
negation	Neg: 1;
word clusters	probably, definitely, def
all-caps	YES, COOL
punctuation	!+: 1, ?+: 0, !?+: 0
emoticons	:D, >:(
hashtags	#excited, #NowPlaying
elongated words	soooo, yaayyy

Improving Sentiment Lexicons

We proposed a lexicon-based approach (Kiritchenko et al., 2014) to determine the sentiment of words in affirmative and negated context.

(1) Hashtags or emoticons are used as the gold labels of sentiment in tweets (Mohammad, 2012).

(2) A tweet corpus is split into two parts: Affirmative Context Corpus and Negated Context Corpus.



(3) For each word *w* in a tweet corpus, an association score is generated using the Affirmative and Negated Context, separately.
score(w) = PMI(*w*, positive) – PMI(*w*, negative)
- If score(*w*) > 0, then word *w* is positive
- If score(*w*) < 0, then word *w* is negative

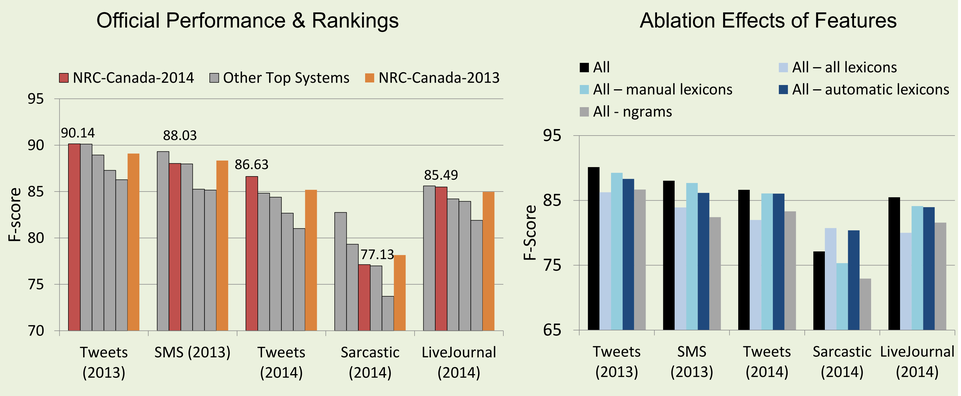
Examples: garbage -0.65 garbage_negatedContext 0.10

(4) Using these new lexicons, we generate sentiment-lexicon features and add them to our systems (e.g., *# sentiment words* and *sentiment scores*.)

Discriminating Negation Words

- Different negation words, e.g., *never* and *didn't*, often have different effects on sentiment (Zhu et al., 2014; Taboada et al., 2011).
- In our term-level system, we take this into consideration. For example, a word *acceptable* is treated as *acceptable_beNever* and *accetptable_cannot* when it appears after word *never* and *cannot*, respectively.

Results: Task A

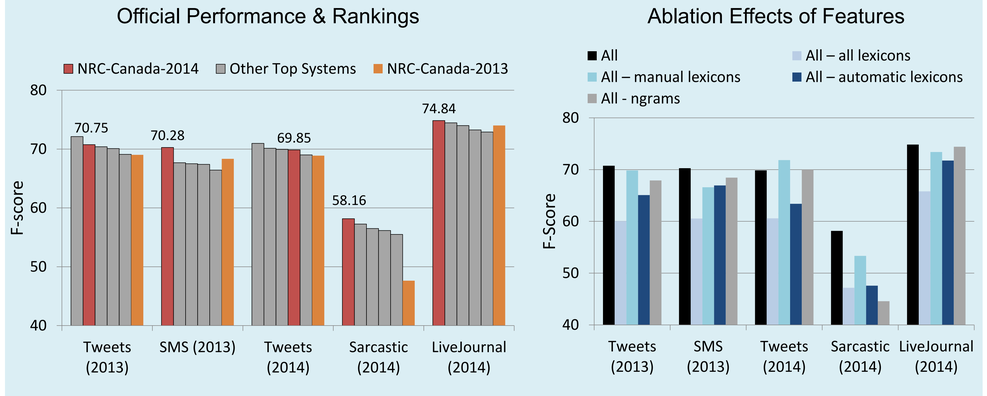


Conclusions

- Best micro-averaged results over all 5 datasets in both tasks.
- Message-level task: best results on 3 out of 5 datasets; term-level task: best results on 2 out of 5 datasets.
- Most useful features: sentiment lexicons, especially automatic tweet-specific lexicons.

NRC Sentiment Lexicons are available for download:
www.purl.com/net/sentimentoftweets

Results: Task B



National Research Council Canada
Conseil national de recherches Canada

Canada