Svetlana Kiritchenko and Saif M. Mohammad
National Research Council Canada
Svetlana.Kiritchenko@nrc-cnrc.gc.ca

NRC·CNRC Information and Communications Technologies

Happy Accident: A Sentiment Composition Lexicon for Opposing Polarity Phrases

In Proceedings of the 10th edition of the Language Resources and Evaluation Conference, May 2016, Portorož (Slovenia)

1. Our Contribution

We created a sentiment lexicon of phrases and their constituent words: a sentiment composition lexicon (SCL). Each phrase includes at least one positive word and at least one negative word—we call such phrases opposing polarity phrases (OPP).

▲happy + ▼accident = ▲happy accident

Annotations:
- done manually with real-valued score of sentiment association using Best–Worst Scaling technique;
- are shown to be reliable.

SCLs, such as SCL-OPP and SCL-NMA [1], are useful in understanding how meaning (and sentiment) is composed.

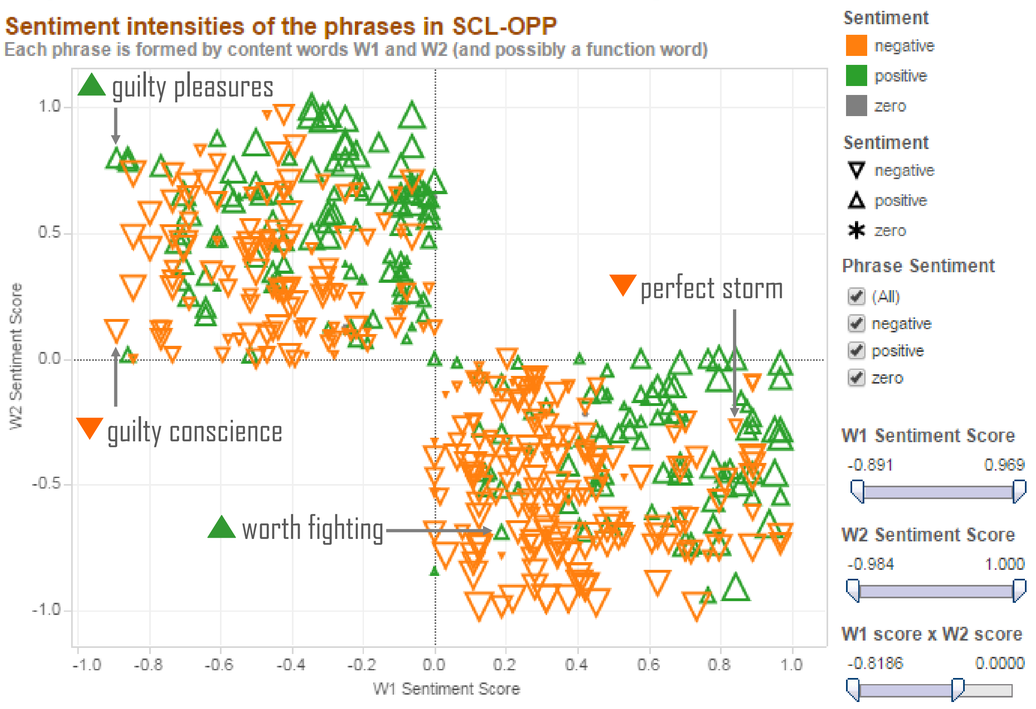
2. SCL-OPP

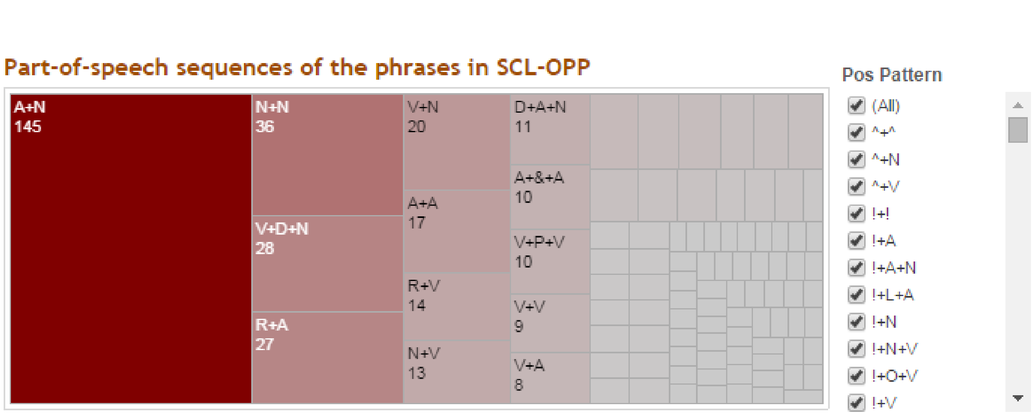
Sentiment Composition Lexicon for Opposing Polarity Phrases (SCL-OPP) includes 1,178 terms: 311 bigrams, 265 trigrams, and 602 single-word constituents.

Example term	Sentiment	Pattern
heart attack	-0.906 ▼	▲noun + ▼noun
long time friends	0.734 ▲	▼adj. + ▲noun + ▲noun
guilty pleasures	0.484 ▲	▼adj. + ▲noun
bad luck	-0.750 ▼	▼adj. + ▲noun
bad	-0.406 ▼	
luck	0.578 ▲	

Interactive Visualization

http://www.saifmohammad.com/WebPages/SCL.html#OPP





The size of a triangle is proportional to the absolute value of the phrase's sentiment score. The size of a tile is proportional to the number of instances corresponding to that pattern.

3. Creating SCL-OPP

Term selection: opposing polarity bigrams and trigrams from tweets; the polarities of the words were determined by look-up in existing sentiment lexicons.

Annotation:
- done with Best–Worst Scaling (Louviere and Woodworth, 1990): a comparative annotation scheme commonly used in marketing research; has been shown to produce reliable annotations of terms for sentiment [2];
- done manually by crowdsourcing;
- each question was annotated by eight respondents.

CrowdFlower

Annotation questions: Given a 4-tuple (4 terms),
- identify the term that is associated with the most amount of negative sentiment;
- identify the term that is associated with the most amount of positive sentiment.

most negative	4-tuple	most positive
○	shameless self promotion	○
○	happy tears	○
○	hug	○
○	major pain	○

Obtaining real-valued scores (Orme, 2009):

$$score(t) = \frac{\#most\ positive(t) - \#most\ negative(t)}{\#annotations(t)}$$

Quality of annotation: the ranking of words by sentiment is remarkably consistent even when the annotation process is repeated with a different set of annotators (Spearman's ρ = 0.98).

4. Applications

- studying sentiment composition [3];
- evaluating automatic methods: SemEval-2016 shared task on determining sentiment intensity of English and Arabic phrases (http://alt.qcri.org/semeval2016/task7/);
- studying how human brain processes sentiment;
- automatically creating a large coverage sentiment lexicon of multi-word phrases.

5. References

[1] Kiritchenko and Mohammad. The Effect of Negators, Modals, and Degree Adverbs on Sentiment Composition. *WASSA-2016*.
[2] Kiritchenko and Mohammad. Capturing Reliable Fine-Grained Sentiment Associations by Crowdsourcing. *NAACL-2016*.
[3] Kiritchenko and Mohammad. Sentiment Composition of Words with Opposing Polarities. *NAACL-2016*.

SCL-OPP is available at:
http://www.saifmohammad.com/WebPages/SCL.html#OPP

Code for Best–Worst Scaling is available at:
http://www.saifmohammad.com/WebPages/BestWorst.html