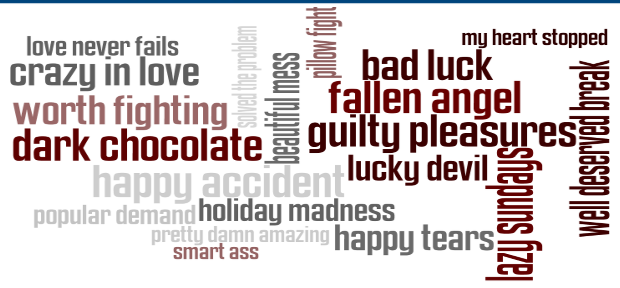
Svetlana Kiritchenko and Saif M. Mohammad
National Research Council Canada
svetlana.kiritchenko@nrc-cnrc.gc.ca, saif.mohammad@nrc-cnrc.gc.ca

NRC·CNRC

Information and Communications Technologies



Sentiment Composition of Words with Opposing Polarities

In Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, June, 2016, San Diego, California.

1. Our Contribution

Sentiment composition is the determining of sentiment of a multi-word linguistic unit (phrase or sentence) based on its constituents. We study sentiment composition in phrases that include at least one positive word and at least one negative word—opposing polarity phrases (OPP).

▲happy + ▼accident = ▲happy accident

We created a sentiment composition lexicon for opposing polarity phrases (SCL-OPP) that provides sentiment associations for OPPs and their constituent words [1].

In this work, we use SCL-OPP to:

- analyze the linguistic patterns in OPPs;
- apply unsupervised and supervised techniques of sentiment composition to determine their efficacy on OPPs.

2. Sentiment Composition Lexicon for Opposing Polarity Phrases (SCL-OPP)

- includes 1,178 terms: 311 bigrams, 265 trigrams, and 602 single-word constituents;
- was annotated with real-valued score of sentiment association using Best–Worst Scaling technique and crowdsourcing;
- includes phrases corresponding to many different Sentiment Composition Patterns, and, therefore, useful for:
  - studying sentiment composition;
  - evaluating sentiment composition algorithms.

3. Automatic Sentiment Composition in OPPs

We investigate whether accurate models of sentiment composition for OPPs can be learned. We conduct experiments with several unsupervised and supervised techniques using features, such as unigrams, parts of speech (POS), sentiment scores, and word embeddings.

4. Conclusions

- The sentiment of an opposing polarity phrase cannot be reliably predicted only from POS and sentiment of the constituents.
- Constituent unigrams, their POS, their sentiment scores, and their embedding vectors are all beneficial in supervised sentiment prediction in OPPs.

5. References

[1] Kiritchenko and Mohammad. Happy Accident: A Sentiment Composition Lexicon for Opposing Polarity Phrases. *LREC-2016*.
[2] Kiritchenko and Mohammad. Capturing Reliable Fine-Grained Sentiment Associations by Crowdsourcing and Best–Worst Scaling. *NAACL-2016*.

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

2 (expanded). Sentiment Composition Lexicon for Opposing Polarity Phrases

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):

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

Sentiment Composition Patterns

Sentiment Composition Pattern (SCP) is a rule:

$$\underbrace{\mathrm{POS}(w_1),\ \mathrm{sentiment}(w_1) + \mathrm{POS}(w_2),\ \mathrm{sentiment}(w_2)}_{\text{left-hand side (LHS)}} \longrightarrow \underbrace{\mathrm{sentiment}(w_1+w_2)}_{\text{right-hand side (RHS)}}$$

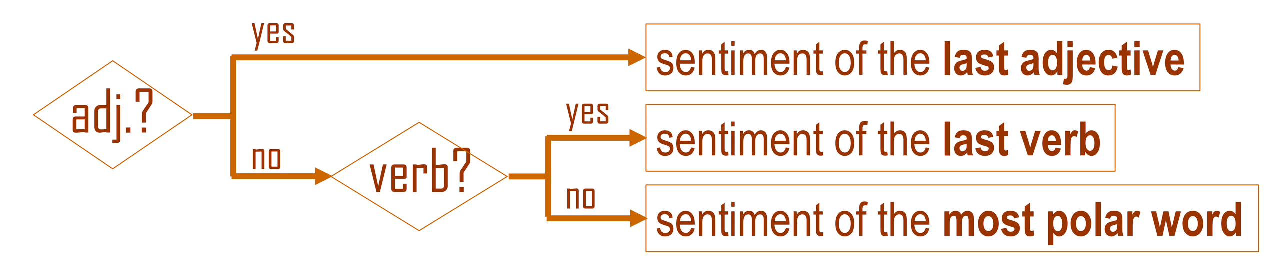
$$Occurrence\ rate\ (SCP) = \frac{freq(LHS \to RHS)}{freq(LHS \to positive) + freq(LHS \to negative)}$$

Example Sentiment Composition Pattern	Occ. rate	# phrases	Example terms
▼adj. + ▲adj. → ▲phrase	0.76	17	guilty pleasures
▼adj. + ▲noun → ▼phrase	0.59	68	bad luck
▲adj. + ▼noun → ▼phrase	0.53	73	perfect storm
▲adv. + ▼adj. → ▼phrase	0.89	18	plain sad
▲noun + ▼noun → ▼phrase	0.52	25	heart attack
▼verb + det. + ▲noun → ▼phrase	0.65	17	lost a child
▼verb + ▲noun → ▼phrase	0.82	17	losing hope

3 (expanded). Automatic Sentiment Composition in Opposing Polarity Phrases

Baseline classifiers

a. Majority label: most frequent polarity label in the dataset
b. Last unigram: sentiment score (label) of the last unigram
c. Most polar unigram: sentiment score (label) of the word with the highest absolute sentiment score
d. POS rule:

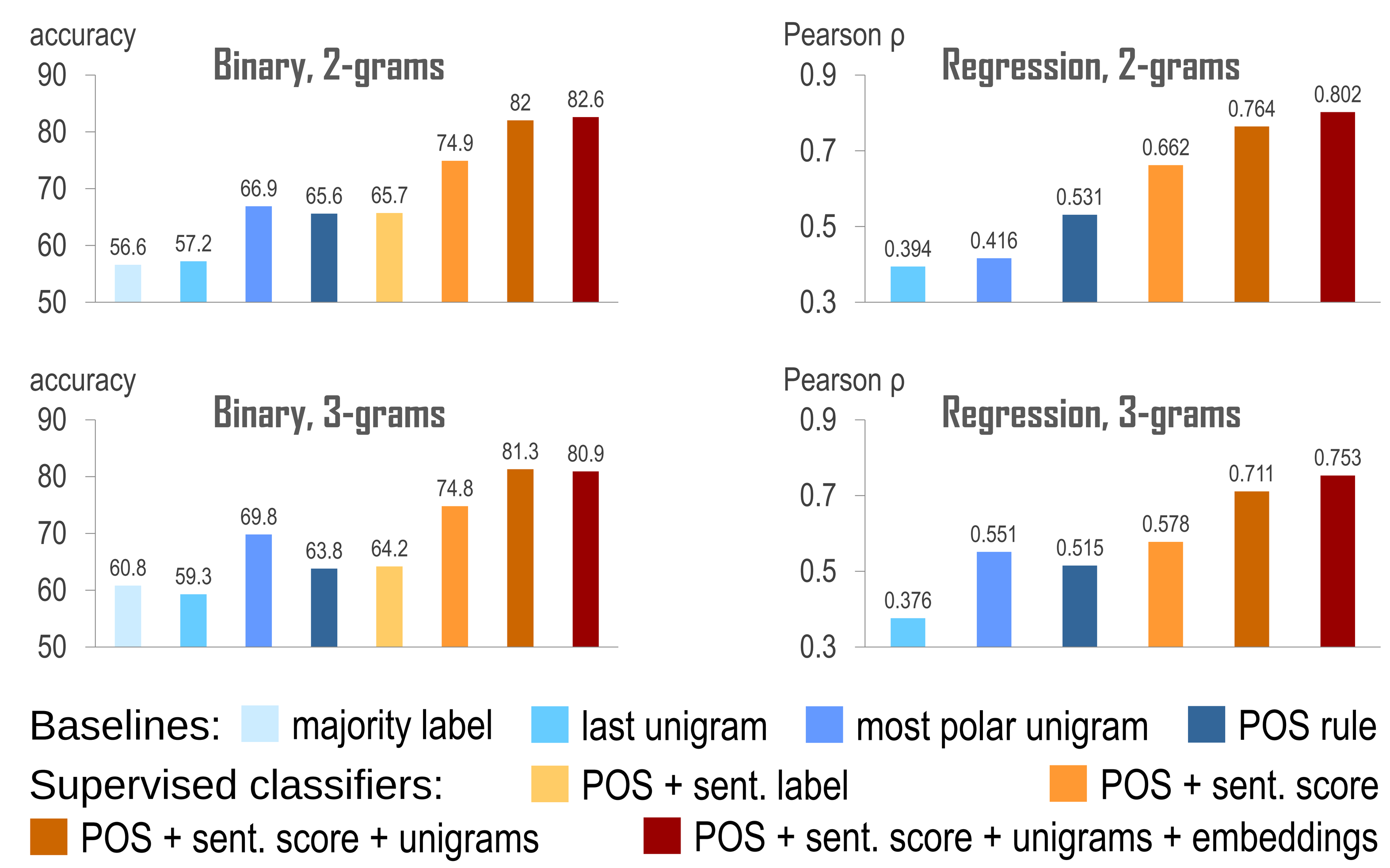


Supervised classifiers

Support Vector Machines (SVM) with RBF kernel; 10-fold cross-validation.

Features:
- unigrams
- POS tags
- sentiment labels
- sentiment scores
- embeddings

Results



Observations

- Sent. of the last unigram is not predictive of the phrase's sent.;
- Adj. and verbs do not always dominate the sent. in a phrase;
- Real-valued sentiment scores of unigrams are substantially more beneficial than binary labels;
- Sentiment of a phrase depends on its constituents and not only on their sentiment;
- Best results are achieved with all the features.

National Research Council Canada | Conseil national de recherches Canada

Canada