Sentiment

William Lowe

Hertie School

5th November 2020

Sentiment analysis

- → Are these documents getting more positive (negative) over time?
- → Do people like this product (party, person)?

My unpopular opinion (which you need not share): Not very interesting, and really a field of its own…

- → An active applied subfield of computer science
- → Some applications in social science
- → Huuuuuge applications in marketing

Sentiment analysis

- → Are these documents getting more positive (negative) over time?
- → Do people like this product (party, person)?

My unpopular opinion (which you need not share): Not very interesting, and really a field of its own…

- → An active applied subfield of computer science
- → Some applications in social science
- → Huuuuuge applications in marketing

Theoretically ambiguous: mostly mixes positive affect, optimism, happiness, etc. and their opposites together into

- → a continuous measure: negative ⟷ positive
- → an ordinal scale, e.g. 0-5 stars
- → a classification, e.g 0 (‘negative’) or 1 (‘positive’)

Measuring sentiment

Consequently we can get sentiment measures in a lot of ways

- → Dictionary models
- → Scaling models
- → Classification models

We'll consider a mix of the first two and the third

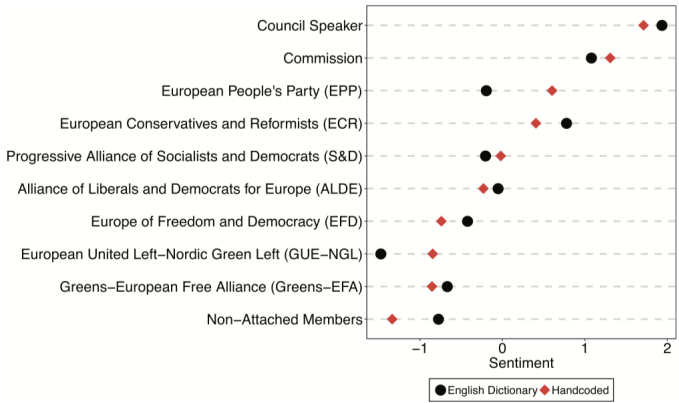
Why a mix?

- → In dictionary applications sentiment is operationalised as the *relative prevalence of positive over negative* terms or mentions
- → First run dictionary to get 'pos' and 'neg' counts, then transform, e.g. as a smoothed logit

$$\log\left(\frac{\text{pos} + \alpha}{\text{neg} + \alpha}\right)$$

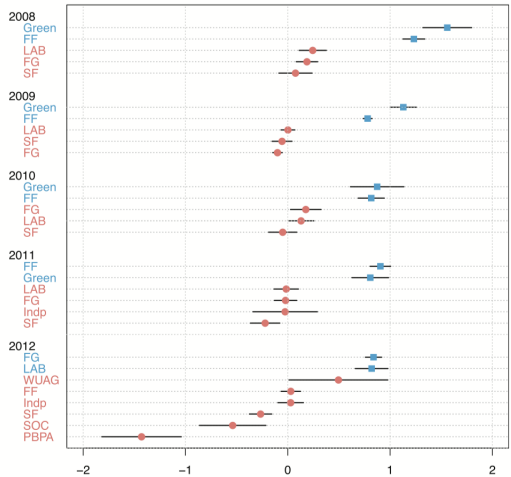
which quanteda calls *polarity*

Dictionaries vs handcoding



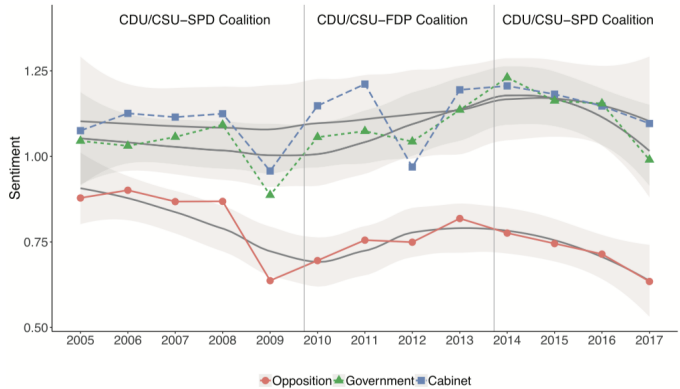
EP State of the Union debate 2010 using (english) Lexicoder dictionaries and polarity measure (Proksch et al., 2019)

Sentiment and government vs opposition



Budget debates in Ireland (Proksch et al., 2019)

Sentiment over time

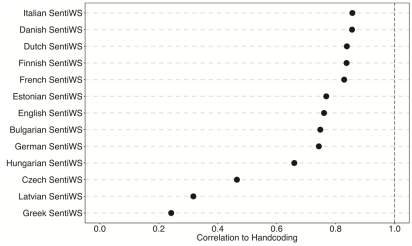


Sentiment on government bills in the German legislature, 2015-2017 (Proksch et al., 2019)

Sentiment as cross-linguistic anchoring

- → Text comparisons across languages is just hard
- → Hand translation is often not available, expensive, and difficult to evaluate
- → Maybe tracking some basic human interaction features is easier to compare?

Proksch et al. (2019) hoped so. I'm not sure if we were right.



Note: there's still translation happening, but of fewer 'easier' terms from a dictionary

Dictionaries

Lots of dictionaries out there. `quanteda.sentiment` (from github) contains several

- → Nielsen (2011) new Affective Norms for English Words (ANEW)
- → Stone et al. (1966) Augmented General Inquirer
- → Hu and Liu (2004) Positive and negative terms
- → Loughran and Mcdonald (2011) Sentiment Word Lists
- → Albugh et al. (2013) Lexicoder Sentiment Dictionary (2015)
- → Mohammad and Turney (2013) NRC Word-Emotion Association Lexicon
- → Rauh (2018) German Political Sentiment Dictionary
- → Remus et al. (2010) 'Senti Wortschatz' (SentiWS)

Classification

Word lists are necessarily domain *un*specific

If we need a sense of sentiment specific to a particular discourse or institution, then

- → Hand construct a dictionary
- → Categorize a set of examples and try to train a classifier

Rather generally

expert constructed dictionary *beats* classifier *beats* general purpose dictionary

but your mileage may vary…

Classification

The easiest way to get a practical feel for how a classifier would work is to make one work

Let's do that next…

References

Albugh, Q., Sevenans, J. & Soroka, S. (2013). *Lexicoder topic dictionaries*.

Hu, M. & Liu, B. (2004). 'Mining and summarizing customer reviews'. 10.

Loughran, T. & Mcdonald, B. (2011). 'When is a liability not a liability? textual analysis, dictionaries, and 10-ks'. *The Journal of Finance*, *66*(1), 35–65.

Mohammad, S. M. & Turney, P. D. (2013, August 28). *Crowdsourcing a word-emotion association lexicon*. Retrieved November 5, 2020, from http://arxiv.org/abs/1308.6297

Nielsen, F. Å. (2011, March 15). *A new anew: Evaluation of a word list for sentiment analysis in microblogs*. Retrieved November 5, 2020, from http://arxiv.org/abs/1103.2903

Proksch, S.-O., Lowe, W., Wäckerle, J. & Soroka, S. (2019). 'Multilingual sentiment analysis: A new approach to measuring conflict in legislative speeches'. *Legislative Studies Quarterly*, *44*(1), 97–131.

References

Rauh, C. (2018). 'Validating a sentiment dictionary for german political language—a workbench note'. *Journal of Information Technology & Politics*, *15*(4), 319–343 _eprint: https://doi.org/10.1080/19331681.2018.1485608.

Remus, R., Quasthoff, U. & Heyer, G. (2010). 'Sentiws – a publicly available german-language resource for sentiment analysis'. *Proceedings of the 7th International Language Resources and Evaluation (LREC'10)*, 1168–1171.

Stone, P. J., Dunphy, D. C., Smith, M. S. & Ogilivie, D. M. (1966). 'The general inquirer'. MIT Press.