

Attitude beyond the word level

Pranav Anand

Department of Linguistics, UCSC

panand@ucsc.edu

The past decade has witnessed increased attention in the semantics of evaluative language across a wide swath of linguistics [3, 9, 7, 2]. Much of the advances have emphasized lexical representations: discovering and curating terms that reliably pick out some granular notion of polarity or emotion. While this approach has achieved notable success in characterizing aggregate sentiment in evaluatively-charged settings (e.g., product or movie reviews), it is still unknown if, in the large, across a wide range of genres, people tend to broadcast their opinions with such conspicuous language. In the present paper, I would like to examine this question from the view of stance classification in online political/argumentative discourse, which I am starting to believe requires building past lexical models of sentiment to make significant gains in reliably classifying writer viewpoint.

Argumentative discourse is a mainstay of human experience, but outside of a few parochial and restricted genres (e.g., philosophical dialogues, staged debates), until very recently, such discussions have occurred in primarily face-to-face interactions. Discussion forums and social media have made these private interactions truly public and at scale, promising the potential for rich, detailed automatically-generated profiles of public mood on a variety of issues. But determining clear stances for individuals from online discourse can be quite tricky, both because such discussions are often at such a fine-grained level that more holistic judgments may not be telegraphed in the text and because people’s particular positions on hot-button issues like abortion or governmental healthcare policy are quite complex, and resist simple two-way or three-way classification. In that regard, the proliferation of online debate sites (e.g., forandagainst.com, convince.net, create-debate.org, etc.) provides a welcome first step into the more wild territory of open-ended discussions. On debate sites, one participant sets a topical question and frames it in terms of a limited set of sides (often two or three); other participants then provide arguments for the side of their choice, as well as rebuttals or supporting arguments. Thus, posts on these sites provide user-generated annotations from a small, fixed set of labels, precisely as in more conventional review datasets.

At UCSC, we have built siding classifiers for several thousand posts on convinceme.net across 14 topics [11]. Although classifier performance has overall been low on this task, one important finding we have uncovered is that features derived from sentiment lexicons (e.g., target-sentiment pairs) do not reliably improve performance above a simple unigram baseline. Given the focus of much sentiment research, this may seem surprising – surely in determining whether abortion should be legal, it would be significant to record observations like “abortion is wrong” or “abortion is a terrible idea.”

I suggest that two factors are at issue. The first is that since much of the content on these sites takes place in conversations, there is frequent repetition, recasting, and attribution of others’ remarks. Hence, extracting features in a very local window may obscure larger, more contextual cues to side. In addition, it appears that much of argumentative discussion actually concerns issues, claims, and points of contention that are expressed in “objective” language. The following conversation from an abortion debate is illustrative of both of these points:

S1: Even if you want a baby, its unbelievably hard to go through and something your body never fully recovers from....You cant force somebody to go through this life-threatening condition that is always physically harmful.

S2: “Life threatening condition that is always physically harmful”? What a giant load of steamy BS. Rarely is pregnancy physically harmful and even rarer is it life threatening.

S1: Pregnancy is ALWAYS physically harmful. You try carrying a load of extra weight about and see what that does to your heart.

S2: So, tell me, if it is ALWAYS life-threatening, why are there so few life-threatening conditions mentioned above?

In this extract, the proximal question under discussion is whether pregnancy should be considered inherently dangerous (life-threatening and harmful). In forming a coherent discourse, they continually pick up on each other’s previous comment, though they indicate their viewpoint of that comment via a host of mechanisms (conditionalization, questions, anaphoric exclamation). Unsurprisingly, then, relatively local cues, like n-grams or relations between n-grams will lead to noisy results. In addition, note that the debate does not invoke any conventional markers of sentiment in a way that would reliably indicate side. Rather, it is only through our understanding of the authors’ opinion on the question of pregnancy’s implicit harm that we as readers come to a conclusion.

It is this last point that I think is worth careful attention for the field as a whole. The idea that sentiment can be conveyed not simply by particular words, but by more complex ‘compositional’ structures is now under systematic study [8, 5, 1, 4, 6, 10]. While as a formal semanticist I find this move aesthetically pleasing, it is worth pointing out that in a real sense much of this nascent literature seeks to replace the *semantics* of lexical sentiment terms with a far more *pragmatic* picture; that is, tracking use and commonsense reasoning in place of a reliable mapping from words to evaluative stance. Under such a picture, for example, ‘life-threatening’ conveys negative sentiment in virtue of the presumed high value of (one’s) life and the fact that threatening something entails possible harm to it; taken together, harming something valuable is tendentially undesirable.

Such systems trade the high-precision of terms like ‘terrible’ or ‘fantastic’ for the higher recall of recognizing expressions like ‘your body never fully recovers from’ as evaluative. But pragmatic reasoning is far more sophisticated, nuanced, defeasible, and culturally-circumscribed than many conventional sentiment terms, resisting the kind of turn-key operation that a sentiment lexicon invites. The central question becomes how much world knowledge we need to build into such systems to reliably extract writer sentiment on a general scale.

With colleagues at the Naval Postgraduate School, I have been examining how governmental policy positions are discussed on Facebook. The following seven posts come from a Facebook page opposing a recent proposal by the Singaporean government to radically increase immigration into the country (in order to ensure the solvency of entitlements). All seven express (in the views of our Singaporean annotators) strong opposition to the proposal.

Suck thumb and buy the cheapest and oldest car u can find.

Haha train broke down we must walk home.

Oh..im sooo looking forward to 2016.

What had happen to us..Our fore fathers are immigrants too!

How does this work? Every employee will get 40% increase of their salary for 3 years?

government pays employers money and the employers pay employees peanuts

what jerry say is true. if the money is given to the employer employer will just give workers market rate and that money they keep

The posts express, respectively, a kind of worst-case resignation, an anecdotal argument based on the current infrastructural inadequacy, a sarcastic comment about the proposal start date, a claim about prior immigrant waves, a claim about the unrealistic expectations of the government plan, and two worries about how governmental funds will trickle down to the average citizen. It is not entirely clear to me how extensible existing theories of compositional sentiment are to capture the sheer richness of expressive content. While ‘train break down’ could be modeled in the vein of ‘life-threatening’, it is hard to see how this model would extend simply to the worries about the employers or buying cheap items without a rich world knowledge to build from.

But I suspect that for the next phase of sentiment detection this is how we will need to proceed, at least for argumentative discussions. In general, I suspect that these more pragmatic sources of sentiment are rife outside of newswire and reviews, and that we are missing a substantial piece of the evaluative pie by ignoring them.

References

- [1] Y. Choi and C. Cardie. Learning with compositional semantics as structural inference for subsentential sentiment analysis. In *Proceedings of EMNLP 2008*, 2008.
- [2] R. Jackendoff. *Language, consciousness, culture*. MIT Press, 2007.
- [3] J. R. Martin and P. R. R. White. *Language of Evaluation: Appraisal in English*. Palgrave Macmillan, 2005.
- [4] K. Moilanen, S. Pulman, and Y. Zhang. Packed feelings and ordered sentiments: Sentiment parsing with quasi-compositional polarity sequencing and compression. In *Proceedings of WASSA 2010, EACI 2010*, 2010.
- [5] T. Nasukawa and J. Yi. Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture*, 2003.
- [6] A. Neviarouskaya, H. Prendinger, , and M. Ishizuka. Recognition of affect, judgment, and appreciation in text. In *Proceedings of COLING 2010*, 2010.
- [7] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135, 2008.

- [8] L. Polanyi and A. Zaenen. Contextual valence shifters. in computing attitude and affect in text. In J. W. James G. Shanahan, Yan Qu, editor, *Computing Attitude and Affect in Text: Theory and Application*. Springer Verlag, Dordrecht, The Netherlands, 2005.
- [9] C. Potts. *The Logic of Conventional Implicature*. Oxford University Press, 2005.
- [10] K. Reschke and P. Anand. Extracting contextual evaluativity. In *Proceedings of ICWS 2011*, 2011.
- [11] M. Walker, P. Anand, J. E. F. Fox Tree, C. Martell, and J. King. That's your evidence?: classifying stance in online political and social debate. *Decision Support Sciences*, 2012.