# Constructive Language in News Comments

Varada Kolhatkar
Discourse Processing Lab
Simon Fraser University
Burnaby, Canada
vkolhatk@sfu.ca

Maite Taboada
Discourse Processing Lab
Simon Fraser University
Burnaby, Canada
mtaboada@sfu.ca

## Abstract

We discuss the characteristics of constructive news comments, and present methods to identify them. First, we define the notion of constructiveness Second, we annotate a corpus for constructiveness. Third, we explore whether available argumentation corpora can be useful to identify constructiveness in news comments. Our model trained on argumentation corpora achieves a top accuracy of 72.59% (baseline = 49.44%) on our crowd-annotated test data. Finally, we examine the relation between constructiveness and toxicity. In our crowd-annotated data, 21.42% of the non-constructive comments and 17.89% of the constructive comments are toxic, suggesting that non-constructive comments are not much more toxic than constructive comments.

## 1 Introduction

The goal of online news comments is to provide constructive, intelligent and informed remarks that are relevant to the article, often in the form of an exchange with other readers. Many comments, however, do not contribute to achieving this goal. Online comments have a broad range: they can be vacuous, dismissive, abusive, hateful, but also constructive. Below we show two comments on an article about Hillary Clinton's loss in the presidential election in 2016.[1]

(1) I have 3 daughters, and I told them that Mrs. Clinton

Proceedings of the First Workshop on Abusive Language Online, pages 11–17,
[illegible]

| Training | Validation accuracy (%) | Test accuracy (%) |
|---|---|---|
| YNC + AEC | 68.43 | 68.45 |
| YNC | 72.76 | 72.59 |
| AEC | 69.30 | 52.54 |

Table 1: Constructiveness prediction results using argumentation corpora. The test data was our annotated constructiveness data in all cases. Random baseline accuracy = 49.44%.

| Feature | OR |
|---|---|
| Argumentative discourse relations | 3.49 |
| Stance adverbials | 2.52 |
| Reasoning verbs & modals | 2.02 |
| Root clauses | 1.37 |
| Conjunctions & connectives | 0.82 |
| Abstract nouns | 0.51 |

Table 2: Association of constructiveness with linguistic features in terms of OR (odds ratio).

We trained with the ADAM stochastic gradient descent for 10 epochs. The important parameter settings are: batch size = 512, embedding size = 200, drop out = 0.5, and learning rate = 0.001.

We wanted to examine which argumentation dataset is more effective in identifying constructiveness. So we carried out experiments with different train and test combinations. In each experiment, 1% of the training data was used as the validation set.

Table 1 shows the average validation and test accuracies for three runs with the same parameter settings. Below we note a few observations. First, we achieved the best result when YNC was included in the training set. Second, AEC seems not to have much effect on the test accuracy but YNC does; when we do not have YNC in the training data, the results drop markedly. This might be because the size of the AEC corpus is relatively small and the model was not able to learn any relevant patterns from this data. Finally, the validation and test accuracy is more or less same for the first two rows, when YNC is included in the training data.

### 3.2 Association with argumentation features

In addition to the classifier described above, we also examine the association between constructiveness and a number of linguistic and discourse features typically found in argumentative texts, based on the extensive literature on argumentation (Biber, 1988; van Eemeren et al., 2007; Moens et al., 2007; Tseronis, 2011; Becker et al., 2016; Habernal and Gurevych, 2017; Azar, 1999; Peldszus and Stede, 2016). We calculate association in terms of odds ratio (Horwitz, 1979), which tells us the odds of a comment being constructive in the presence of a feature. Results are shown in Table 2. We observed a strong association between constructiveness and occurrence of argumentative discourse relations (Cause, Comparison, Condition, Contrast, Evaluation and Explanation[10]). The odds ratio for argumentative discourse relations is 3.49, which means that constructive texts are 3.49 times more likely to have this feature than non-constructive texts. Other features with strong association with constructiveness are stance adverbials (e.g., *undoubtedly, paradoxically, of course*), and reasoning verbs (e.g., *cause, lead*) and modals. Root clauses (clauses with a matrix verb and an embedded clause, such as *I think that ...*) show a medium association with constructiveness. On the other hand, abstract nouns (e.g., *issue, reason*) and, surprisingly, conjunctions and connectives are not associated with constructive texts. The latter is surprising because many discourse relations contain a connnective.

## 4 Toxicity in news comments

In the context of filtering news comments, we are also interested in the relationship between constructiveness and toxicity. We propose the label toxicity for a range of phenomena, including verbal abuse, offensive comments and hate speech. To better understand the nature of toxicity and se-

| | C (n = 603) | Non-C (n = 518) |
|---|---|---|
| Not toxic | 82.09% | 78.57% |
| Mildly toxic | 16.08% | 15.44% |
| Toxic | 1.33% | 5.21% |
| Very toxic | 0.50% | 0.77% |
| Total | 100% | 100% |

Table 3: Percent distribution of constructive and toxic comments in CrowdFlower annotation. C = Constructive.

comments were described as those which may be considered toxic only by some people, or which expressed anger and frustray beNcLfby disb27ion2sideredity(on2sid [Constructls20(27iby(on2si59(of)-3-rustr-a

structure: An application of Rhetorical Structure Theory. Argumentation 13(1):97–144.

Maria Becker, Alexis Palmer, and Anette Frank. 2016. Clause types and modality in argumentative microtexts. In Proceedings of the Workshop on Foundations of the Language of Argumentation (in conjunction with COMMA 2016). Postdam, pages 1–9.

Douglas Biber. 1988. Variation across Speech and Writing. Cambridge University Press, Cambridge.

Dirk Brand and Brink Van Der Merwe. 2014. Comment classi cation for an online news domain. In Proceedings of the First International Conference on the use of Mobile Informations and Communication Technology in Africa UMICTA. Stellenbosch, South Africa, pages 50–56.

Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber. 2017. Automated hate speech detection and the problem of offensive language. In Proceedings of the 11th International Conference on Web and Social Media. Montréal.

Nicholas Diakopoulos. 2015. Picking the NYT Picks: Editorial criteria and automation in the curation of online news comments. ISOJ Journal 6(1):147–166.

Alex Graves and Jürgen Schmidhuber. 2005. Framewise phoneme classi cation with bidirectional LSTM networks. In Proceedings of the IEEE International Joint Conference on Neural Networks, IJCNN. volume 4, pages 2047–2052.

Ivan Habernal and Iryna Gurevych. 2017. Argumentation mining in user-generated web discourse. Computational Linguistics 43(1):125–179.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. Neural Computation 9(8):1735–1780.

Ralph I. Horwitz. 1979. A method of estimating comparative rates from clinical data: Applications to cancer of the lung, breast, and cervix: Corn eld J: J Nat Cancer Inst 11: 12691275, 1951. Journal of Chronic Diseases 32(1-2):i.

Abhyuday N. Jagannatha and Hong Yu. 2016. Bidirectional RNN for medical event detection in electronic health records. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego, CA, pages 473–482.

Sha q Joty, Giuseppe Carenini, and Raymond Ng. 2015. CODRA: A novel discriminative framework for rhetorical analysis. Computational Linguistics 41(3):385–435.

Irene Kwok and Yuzhou Wang. 2013. Locate the hate: Detecting tweets against blacks. In Proceedings of the Twenty-Seventh AAAI Conference on Arti cial Intelligence. AAAI'13, pages 1621–1622.

Marie-Francine Moens, Erik Boiy, Raquel Mochales Palau, and Chris Reed. 2007. Automatic detection of arguments in legal texts. In Proceedings of the 11th international conference on Arti cial intelligence and law. ACM, Stanford, California, pages 225–230.

Elaheh Momeni, Claire Cardie, and Nicholas Diakopoulos. 2015. A survey on assessment and ranking methodologies for user-generated content on the web. ACM Computing Surveys 48(3):1–49.

Courtney Napoles, Joel Tetreault, Aasish Pappu, Enrica Rosato, and Brian Provenzale. 2017. Finding good conversations online: The Yahoo News Annotated Comments Corpus. In Proceedings of the 11th Linguistic Annotation Workshop, EACL. Valencia, pages 13–23.

Vlad Niculae and Cristian Danescu-Niculescu-Mizil. 2016. Conversational markers of constructive discussions. In Proceedings ocee2 -10.959 Td [(good)-408(con)40(

Frans H. van Eemeren, Peter Houtlosser, and A. Francisca Snoeck Henkemans. 2007. *Argumentative Indicators in Discourse: A pragma-dialectical study*. Springer, Berlin.

Zeerak Waseem and Dirk Hovy. 2016. Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. San Diego, CA, pages 88–93.

Ellery Wulczyn, Nithum Thain, and Lucas Dixon. 2016. Ex machina: Personal attacks seen at scale. arXiv:1702.08138v1