A Corpus for the Analysis of Online 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

There is a growing interest among journalists, social scientists and computational linguists in studying online interactions and mediating them to improve their quality. We present the SFU Opinion and Comments Corpus which can be used to study online interactions in a systematic way. The corpus comprises 10,339 opinion articles from the Canadian national newspaper *The Globe and Mail* and 663,173 comments posted in response to those articles. We also present a description of four layers of annotations carried out on a subset of the corpus: constructiveness, toxicity, negation, and Appraisal.

1 Introduction

Online commenting allows for direct communication among people and organizations from diverse socioeconomic classes and backgrounds on important issues. Popular news articles receive

Constructive comments	Non-constructive comments
target specific points and provide appropriate evidence	present opinions without support
offer a solution to the issues discussed in the article	dismiss the terms of debate
share a related personal story or experience	are provocative or excessively flattering
encourage other readers to participate in the discussion	irrelevant or unsubstantial

Table 1: Some prototypical characteristics of constructive comments

2 SFU Opinion and Comments Corpus

We present the SFU Opinion and Comments Corpus (SOCC), a collection of opinion articles and the comments posted in response to the articles. The articles include all the opinion pieces published in the Canadian newspaper *The Globe and Mail* in the five-year period between 2012 and 2016, a total of 10,339 articles and 663,173 comments from 303,665 comment threads. The corpus contributes by providing a pairing of articles and comments, and introducing the largest dataset of this kind to date. Furthermore, the articles are all opinion articles, not hard news articles. This is important, because it allows for comparisons of evaluative language in both text types, opinion articles and reader comments. Opinion articles are generally subjective and evaluative, but their language tends to be more formal and argumentative. The comments are also subjective; they, however, tend to be more informal and personal in nature. The corpus has been collected with attention to preserving reply structures and other metadata. We have made this corpus publicly available for non-commercial use through GitHub.[3]

3 Annotations

We have been carrying out annotations and analyses on a subset of SOCC (Kolhatkar and Taboada, 2017a,b). In this section, we summarize the results of four annotation experiments.

Constructiveness. There is a growing interest in automatically organizing reader comments in a sensible way (Napoles et al., 2017; Llewellyn et al., 2014). One useful way to organize comments is based on their *constructiveness*, i.e., by identifying which comments provide insight and encourage a healthy discussion. For instance, *The New York Times* manually selects and highlights comments representing a range of diverse views, referred to as *NYT Picks*.

To understand constructiveness in online comments, we annotated a subset of SOCC containing 1,121 comments for constructiveness using crowdsourcing.[4] We define constructiveness in terms of prototypical characteristics which we obtained from a crowdsourced survey, shown in Table 1.

Each comment was annotated by at least three annotators. As we were interested in the verdict of native speakers of English, we limited the allowed demographic region to English-speaking countries. The percentage agreement on a random sample of 100 annotations was 87.88%, suggesting that constructiveness can be reliably annotated. Among 1,121 comments, 603 comments (53.79%) were classified as constructive, 517 (46.12%) as non-constructive, and the annotators were not sure in only one case. To examine the quality of the crowd annotations we asked a moderator to evaluate the acceptability of the crowds answers. For that, we randomly selected 222 instances from the crowd-annotated data and asked the expert whether they agree with the crowds answer or not. Overall the expert agreed with the crowd 77.93% of the time. We carefully curated the crowd annotated corpus, removing duplicates and instances with very low agreement. Our curated corpus contains 1,043 instances and is available for non-commercial use via GitHub.[5]

Toxicity. We are interested in exploring the relationship between *constructiveness* and *toxicity* in reader comments. For that purpose, we added a layer of toxicity annotations on the same subset of 1,121 comments above. We defined four levels of toxicity: very toxic, toxic, mildly toxic, and not toxic. In our annotation guidelines we provide prototypical characteristics of toxicity, as shown in Table 2. Again we used crowdsourcing for annotation and each comment was annotated by at least three annotators.

Among all comments, 203 (19.46%) comments

[3] https://github.com/sfu-discourse-lab/SOCC

[4] https://www.figure-eight.com

[5] https://github.com/sfu-discourse-lab/SOCC#constructiveness

Toxicity level	Prototypical characteristics
Level 4	contain harsh, abusive or offensive language are in ammatory contain a personal attack, insult or condemnation
Level 3	ridicule, tease, or poke others cause embarrassment or disrespect disagreeing aggressively or joking inappropriately
Level 2	express frustration and anger likely to be perceived as toxic by some people in some contexts
Level 1	unlikely to be perceived as toxic

Table 2: Some prototypical characteristics of toxic comments

had some toxicity in them. In our data, we did not nd any signi cant differences in toxicity levels between constructive and non-constructive comments, i.e., constructive comments were as likely

Varada Kolhatkar and Maite Taboada. 2017a. Constructive language in news comments. In *Proceedings of the First Workshop on Abusive Language Online*, pages 11–17, Vancouver.

Varada Kolhatkar and Maite Taboada. 2017b. Using New York Times Picks to identify constructive comments. In *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism*, pages 100–105, Copenhagen.

Clare Llewellyn, Claire Grover, and Jon Oberlander. 2014. Summarizing Newspaper Comments. In *Proceedings of ICWSM*, Ann Arbor, MI.

James R. Martin and Peter R. R. White. 2005. *The Language of Evaluation*. Palgrave, New York.

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*, pages 13–23, Valencia.