

VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text

C.J. Hutto

Eric Gilbert

Georgia Institute of Technology, Atlanta, GA 30032
 cijhutto@gatech.edu gilbert@cc.gatech.edu

gilbert@cc.gatech.edu

Abstract

The inherent nature of social media content poses serious challenges to practical applications of sentiment analysis. We present VADER, a simple rule-based model for general sentiment analysis, and compare its effectiveness to eleven typical state-of-practice benchmarks including LIWC, ANEW, the General Inquirer, SentiWordNet, and machine learning oriented techniques relying on Naive Bayes, Maximum Entropy, and Support Vector Machine (SVM) algorithms. Using a combination of qualitative and quantitative methods, we first construct and empirically validate a *gold-standard* list of lexical features (along with their associated sentiment intensity measures) which are specifically attuned to sentiment in microblog-like contexts. We then combine these lexical features with consideration for five general rules that embody grammatical and syntactical conventions for expressing and emphasizing sentiment intensity. Interestingly, using our parsimonious rule-based model to assess the sentiment of tweets, we find that VADER outperforms individual human raters (*F1 Classification Accuracy* = 0.96 and 0.84, respectively), and generalizes more favorably across contexts than any of our benchmarks.

1. Introduction

Sentiment analysis is useful to a wide range of problems that are of interest to human-computer interaction practitioners and researchers, as well as those from fields such as sociology, marketing and advertising, psychology, economics, and political science. The inherent nature of microblog content - such as those observed on Twitter and Facebook - poses serious challenges to practical applications of sentiment analysis. Some of these challenges stem from the sheer rate and volume of user generated social content, combined with the contextual sparseness resulting from shortness of the text and a tendency to use abbreviat-

zales, & Booth, 2007). Sociolinguists, and computer scientists, cause it has been extensively forwarded dictionary and simplified, understood, and extended to make LIWC an attractive option as a reliable lexicon to extract emotion from text. Despite their pervasiveness in social media contexts, these tools have little regard for their actual su-

This paper describes the evaluation of VADER (for Valence Aware Dictionary and sEntiment Reasoning). We use both qualitative and quantitative methods to validate, a *gold-standard* sentiment lexicon, and compare it to a manually attuned to microblog-like text. We show that these lexical features with customizable rules that embody grammatical conventions that humans use when expressing sentiment *intensity*. We find that the VADER sentiment classification algorithm improves the accuracy of sentiment classification across several domain corpora: New York Times editorials, movie reviews, and tweets. Interestingly, the VADER lexicon performs well in the social media domain. A human evaluation shows that VADER ($r = 0.84$) is as good as individual human raters ($r = 0.84$) in identifying the truth (aggregated group mean) of the sentiment intensity of each tweet. To further inspect the classification performance, we show that VADER ($F1 = 0.96$) actually outperforms human raters ($F1 = 0.84$) at classifying tweets into positive, negative, and neutral categories.