The Challenge of Fake News: Automated Stance Detection via NLP

Evan T. R. Rosenman
Department of Statistics
Stanford University
Stanford, CA 94305
rosenman@stanford.edu

Jeff T. Sheng
Department of Sociology
Stanford University
Stanford, CA 94305
jtsheng@stanford.edu

Abstract

Fake news, or the online sharing of factually incorrect articles, has emerged as a consequence of the widespread adoption of social media as a way to pass news to others. This paper proposes that deep learning and natural language processing can help solve this problem, specifically by exploring a stance detection task between an article's headline and body. The authors first use the Fake News Challenge dataset of approximately 50,000 headline-article pairs to determine the best deep learning model to classify the headline-article relationship as "unrelated," "agrees," "disagrees," or "discusses." Results from over 50 different models showed that a relatively simple "Siamese" Bidirectional LSTM, combined with dropout regularization, gave the highest average F_1 score. This model also outperformed a reference model provided by the Fake News Challenge organizers. The best model is then used in experiments with two original datasets, one of 12,999 "fake news" articles gathered between October 26, 2016 and November 25, 2016 from unreliable sources; the other a matching dataset of about 15,000 "real news" articles published in five reputable sources over the same time period. A novel fake news detection algorithm is deployed, wherein both real and fake news articles are paired with real news headlines on similar topics. Stance detection is applied to the pairings, and a logistic regression classifier is trained to predict whether an article is real, based on these predicted class distributions. The resultant classifier has several highly statistically significant coefficients. The findings illustrate the potential for using stance detection and other NLP methods to combat fake news.

1   Note

Please contact the authors at their email addresses above for more information about this paper.