

Using Bidirectional Encoder Representations from Transformers to Extract Topic-Independent Sentiment Features for Social Media Bot Detection

Authors : Maryam Heidari, James H. Jones Jr.

Abstract : Millions of online posts about different topics and products are shared on popular social media platforms. One use of this content is to provide crowd-sourced information about a specific topic, event or product. However, this use raises an important question: what percentage of information available through these services is trustworthy? In particular, might some of this information be generated by a machine, i.e., a bot, instead of a human? Bots can be, and often are, purposely designed to generate enough volume to skew an apparent trend or position on a topic, yet the consumer of such content cannot easily distinguish a bot post from a human post. In this paper, we introduce a model for social media bot detection which uses Bidirectional Encoder Representations from Transformers (Google Bert) for sentiment classification of tweets to identify topic-independent features. Our use of a Natural Language Processing approach to derive topic-independent features for our new bot detection model distinguishes this work from previous bot detection models. We achieve 94% accuracy classifying the contents of data as generated by a bot or a human, where the most accurate prior work achieved accuracy of 92%.

Keywords : bot detection, natural language processing, neural network, social media

Conference Title : ICOSNA 2020 : International Conference on Online Social Networks and Applications

Conference Location : Lisbon, Portugal

Conference Dates : September 16-17, 2020