Twinsight: A Novel Multifactor Behavioral Analysis Algorithm for Social Media

McInnis, Daniel

The growing popularity of social media has made it possible for computers to easily analyze human communication to identify trends and sentiment; however, limited work has been done to analyze the behavioral patterns of individual users. Developments in this area could be very important in automatic depression detection given that 17% of Americans experience depression at some point in their life but many never seek professional help. Tw-Lexa is a novel algorithm designed to analyze Twitter profiles for symptoms of depression. The algorithm extracts a set of 81 independent features including profile information, parts of speech patterns, interactions with other users, sentiment, picture analysis, and phrase mood connotation. These features vary in format from numerical values, to lists, to entire language models. Once the features are extracted for a group of users, the algorithm then calculates a similarity score between each pair of users for every independent factor. An Artificial Neural Network model is applied on the principal components of the normalized score matrices to separate depressed and happy (opposite of depressed) people into clusters, each with its own set of latent characteristics. In the experimental design, the algorithm accurately constructed clusters of depressed and happy people misclassifying only 2% (assigning them into clusters with opposites). This supported the initial assumption that data from people's Twitter profiles contains information that can be used to identify various types of depression. This algorithm integrated in Twitter could not only be used to detect early signs of depression in people but also analyze for various other behavioral traits allowing for applications in marketing, recruitment, socio-politics, and finance.

Awards Won:
Fourth Award of $500
European Organization for Nuclear Research-CERN: All expense paid trip to tour CERN