Composite Approach to Extremism and Terrorism Web Content Classification

Authors : Kolade Olawande Owoeye, George Weir

Abstract : Terrorism and extremism activities on the internet are becoming the most significant threats to national security because of their potential dangers. In response to this challenge, law enforcement and security authorities are actively implementing comprehensive measures by countering the use of the internet for terrorism. To achieve the measures, there is need for intelligence gathering via the internet. This includes real-time monitoring of potential websites that are used for recruitment and information dissemination among other operations by extremist groups. However, with billions of active webpages, real-time monitoring of all webpages become almost impossible. To narrow down the search domain, there is a need for efficient webpage classification techniques. This research proposed a new approach tagged: SentiPosit-based method. SentiPosit-based method combines features of the Posit-based method and the Sentistrenght-based method for classification of terrorism and extremism webpages. The experiment was carried out on 7500 webpages obtained through TENE-webcrawler by International Cyber Crime Research Centre (ICCRC). The webpages were manually grouped into three classes which include the 'pro-extremist', 'anti-extremist' and 'neutral' with 2500 webpages in each category. A supervised learning algorithm is then applied on the classified dataset in order to build the model. Results obtained was compared with existing classification method using the prediction accuracy and runtime. It was observed that our proposed hybrid approach produced a better classification accuracy compared to existing approaches within a reasonable runtime.