Radical Ideology Mining in Arabic Tweets

Alsaeedi Riyadh
Wasit University, Iraq

Abstract:

Nowadays, the internet and social media platforms are being misused by extremists and terrorists to spread their propaganda, disseminate their messages, and recruit new members. Arabic is the primary language used by extremist Islamists. There is a significant body of research on colloquial English. Unfortunately, there is little work on Arabic text processing for extracting the main idea. Automated processing of Arabic dialects is challenging due to the lack of orthographic standards and the scarcity of annotated data and public resources. Compared to the numerous publications, resources, and dictionaries available in English, extracting the main idea from Arabic texts remains immature. It has fewer publications and very few resources. The lack of resources in terms of studies conducted on detecting extremism in Islamic networks, the inefficient nature of the Arabic language, the diverse sources of ambiguity, and the use of rich metaphorical texts remain some of the most challenging problems facing Arabic NLP researchers. Another challenge facing this research is the limited availability of data. The dataset of 40,000 Arabic tweets that we present in this research has been carefully tagged and filtered to include both radical and non-radical tweets. Machine Learning (ML) was the method we utilized to automate the process of identifying tweets that contain extremism. 20,004 test samples were used to evaluate the Support Vector Machine (SVM) with RBF kernel after it was trained on TF-IDF features. The Accuracy was 91%.

Keywords:

Online social networks, Machine Learning, Aggressiveness and radicalism, Natural Language Processing (NLP), Religious hate in Arabic.