D21 Detecting Linguistic Markers of Religious Extremism in an Online Environment: A Pakistan Case Study

Mariam Dar, PhD, The Institute for Linguistic Evidence, London, ON N6M 0B5, CANADA*

Learning Overview: After attending this presentation, attendees will better understand the role of the internet, social media, and sermons in radicalization of youth in Pakistan, and of forensic computational linguistic analysis for detecting linguistic markers of religious extremism.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by providing multilingual English-Arabic-Urdu text analysis for detecting linguistic markers of extremism for proactive investigations.

The phenomenon of "extremism," "religious terrorism," or "violent radicalization" has grown in recent years, associated with several ideologies. Pakistan is among the top five on the list of countries affected by terrorism/violent extremism.[1] The war on terrorism has killed nearly five million people in Pakistan, Afghanistan, and Iraq combined, resulting in displacement of families, as well as the clear emotional and mental issues arising from such turmoil.

Social media has been playing a vital role in propagating extremism through radicalization of youth. The advantages of using the internet and social media include communication channels that are not bound to national jurisdictions and that are informal, large groups, cheap, decentralized, and anonymous.[2,3] These channels are used in several ways to attract a young audience, publish violent content based mostly on false information, and persuade/recruit youth for violent acts of terrorism. It has become difficult to differentiate between a religious website that publishes only for the sake of awareness/knowledge and pages/websites where exaggerated/made-up religious stories are posted for a sinister agenda. The current pilot focuses on Pakistan as it is among the top ten internet users in Asia with manifest online problems.[4]

Specialized software can play an important part in identifying such platforms and tracking down terrorist organizations by *detecting warning signals/threats* associated with such activities. Machine learning techniques can be used effectively to detect "weak signals," "digital traces" of "linguistic markers" that characterize the warning signals/threats associated with terrorism or religious extremism. The warning behaviors that have the highest potential to be discovered in text and speech content are leakage (the communication of intent to do harm to a third party), fixation (increasing perseveration on the object of fixation), and identification (a desire to identify oneself with previous attackers or a terrorist organization).[5-9] Previous examinations of warning signals have been developed from forensic psychology and behavioral sciences, sociology, and computer science.[3,6,7,9,12,17] Based on previous studies, a definition of "extremist ideology" as an invocation to violence against specific groups justified by an ideological position (e.g., religious affiliation, racial superiority, and other extremist ideas) is used. "Extremist language" is then defined as the invocation through language.

The examinations that actually consider language, however, focus on content analysis, sentiment analysis, critical discourse analysis, and basic corpus linguistics.[14-16,18] Rigorous theoretical linguistics has generally been missing from the discussion.

This presentation focuses on developing objective, operational definitions of "leakage," "fixation," and "identification" from the perspective of formal linguistics and computational linguistics. Formal linguistics includes both semantic and syntactic theory, and computational linguistics uses algorithms for syntactic, semantic, discursive, and orthographic textual analysis. These methods and techniques can be used for any kind of language-based analysis, but this pilot study focuses specifically on data related to Pakistani radicalization. The pilot for methodological development focuses on detecting the warning signals/threats through linguistic markers of religious extremism in an online environment, with linguistic markers defined from linguistic theory and computational algorithms. Data are collected from social networks (Facebook®, Twitter®), websites publishing religious content, which have more than 10,000 followers/subscribers in active discussions.

Data are analyzed using standard methods in linguistics implemented in the Automated Linguistic Identification & Assessment System (ALIAS).[10,11] First, in line with previous studies, ALIAS is used to calculate quantitative rates for words related and unrelated to extremist ideology. This helps to determine which key words or phrases show up repeatedly and provides a baseline of expected terminology of extremism in the social environment. Second, ALIAS is used to perform syntactic and semantic analysis of passages related and unrelated to extremist ideology to determine if there is a correlation between the content (extremist thoughts/ideas/messages) and syntax (what's the syntax in extremist vs. non-extremist phrases/sentences). The methodologies based in linguistics offer another tool for identifying radicalization through language, focusing on both salient and sophisticated features of language.

Reference(s):

1. Smith, B., Gruenewald, J., Roberts, P., and Damphousse, K. (2015). The Emergence of Lone Wolf Terrorism: Patterns of Behavior and Implications for Intervention. *Sociology of Crime, Law and Deviance*, 20, 89-110.
2. Hale, W.(2012). Extremism on the World Wide Web: A research review. Criminal Justice Studies: *A Critical Journal of Crime, Law & Society*, 25(4), 343-356.
3. Neumann, P. (2013). Options and strategies for countering online radicalization in the United States. *Studies in Conflict & Terrorism*, 36(6), 431-459.
4. National Consortium for the Study of Terrorism and Responses to Terrorism (START): Annex of Statistical Information (2015). *Country reports on terrorism*. Retrieved February 26th from The National Bureau Of Counterterrorism and Countering Violent Extremism: https://www.state.gov/documents/organization/257738.pdf.

5. Meloy, J. R. (2011). Approaching and attacking public figures: A contemporary analysis of communications and behaviour. In C. Chauvin (Ed.), *Threatening communications and behaviour: Perspectives on the pursuit of public figures* (pp. 75–101). Washington, DC: The National Academies Press.
6. Meloy, R., Hoffmann, J., Guldimann, A., and James, D. (2012). The role of warning behaviors in threat assessment: An exploration and suggested typology. *Behavioral Sciences & the Law*, 30(3), 256–279.
7. Meloy, R., Hoffmann, J., Roshdi, K., and Guldimann, A. (2014). Some warning behaviors discriminate between school shooters and other students of concern. *Journal of Threat Assessment and Management*, 1(3), 203–211.
8. Meloy, R., and O'Toole, M. (2011). The concept of leakage in threat assessment. *Behavioral Sciences & the Law*, 29(4), 513–527.
9. Cohen, K., Johansson, F., Kaati, L., and Mork, J. (2014). Detecting linguistic markers for radical violence in social media. *Terrorism and Political Violence*, 26(1), 246–256.
10. Chaski, C. (2013). Best Practices and Admissibility of Forensic Author Identification. *Journal of Law and Policy*, 21(2), 333-376.
11. Chaski, C.E. 2005. Who's At the Keyboard? Recent Results in Authorship Attribution. *International Journal of Digital Evidence*. Volume 4:1. Spring 2005. *Available at http://www.ijde.org.*
12. Scarcella A, Page R, Furtado V (2016) Terrorism, Radicalisation, Extremism, Authoritarianism and Fundamentalism: A Systematic Review of the Quality and Psychometric Properties of Assessments. *PLoS ONE* 11(12): e0166947. https://doi.org/10.1371/journal.pone.0166947.
13. Mashechkin, I.V., Petrovskiy, M.I., Tsarev, D.V. et al. Machine Learning Methods for Detecting and Monitoring Extremist Information on the Internet *Programming and Computer Software* (2019) 45: 99. https://doi.org/10.1134/S0361768819030058.
14. Ahmad, S., Asghar, M.Z., Alotaibi, F.M., and Awan, I. Detection and classification of social-media-based extremist affiliations using sentiment analysis techniques. *Human-Centric Computing and Information Sciences* (2019) 9:24 https://doi.org/10.1186/s13673-019-0185-6.
15. Prentice, S., Rayson, P., and Taylor, P.J. The language of Islamic extremism: Towards an automated identification of beliefs, motivations and justifications. *International Journal of Corpus Linguistics* (2012), 17:2, p. 259-286 https://doi.org/10.1075/ijcl.17.2.05pre.
16. Pennebaker, J.W. and Chung, C.K. (2009). Computerized Text Analysis of Al-Qaeda Transcripts. In Krippendork, K. *The Content Analysis Reader*. Los Angeles: Sage.
17. Lucas, B. (2014) Methods for monitoring and mapping online hate speech. *GSDRC Helpdesk Research Report no. 1121*. University of Birmingham.
18. Törnberg, A. and . Törnberg, P. (2016). Muslims in social media discourse: combining topic modeling and critical discourse analysis. *Discourse, Context and Media* 13, p. 132-142. https://doi.org/10.1016/j.dcm.2016.04.003.

Religious Extremism, Linguistic Markers, Religious Extremism in Pakistan