A Survey Over Violent Extremist Detection in Social Media Websites

[1]Aditi Saraf Jain, [2]Dr. Shikha Agarwal, [3]Dr. Jitendra Agarwal
[1,2,3]Rajiv Gandhi Proudyogiki Vishwavidhyalaya, Bhopal, India

Abstract
In current scenario internet become as a de-facto medium of the communication to others. Internet provides an easy and efficient way of communication. There is a huge amount of content flooded over the internet which contains rational and extremist data. That data spread into various social media websites and social media communities to recruit people for ration and violent extremist activities. Contents like white hated music creates hated perception in the people and approach them to work for any terrorist organization or promote nationalism. A review over the techniques which are used to detect and forecast such activities is presented in this paper. Topic models like LDA (Latent Dirichlet Allocation) and CTM (Correlated Topic Model) etc. are used to predict such activities and ARIMA or some other models are used to forecast such activities. A new topic model called DCNT (Doubly Correlated Nonparametric Topic Model) and a multiple word matching or n-gram matching technique is proposed to provide better performance for detecting and predicting such violent extremist activities.

Keywords
LDA (Latent Dirichlet Allocation), CTM (Correlated Topic Model), DCNT (Doubly Correlated Topic Model), Violent Extremist.

I. Introduction

Now-days internet provides an easy and efficient medium of communication, it crosses the borders of the country and anyone can communicate to the other in a fraction of a second. There is a huge amount of content spread over the social media on a daily basis. That contains radical, violent extremist or terrorist organization related content. Such violent extremist content used to influence people to involve various terrorist related activities.

Fig. 1: Cyber Terrorism

A brief description of such activities is shown in Figure 1.1, in that a terrorism content is spread over the internet which used to approach people to join various terrorist activities.

A. Violent Extremism [12]

Violent extremism is the radical perspective of the people and putting that into the violent actions. In such action people wanted changes in the society to fulfil their radical expectations. These views are influenced by the radical organizations, political organization. There can be a war of ideology among these organizations.

B. Factors Influencing Violent Extremism [12]

These are some factors influencing violent extremism:

1. Political Expectations

Various political parties or organizations are spread hate and radical content or view among the people to fulfil their political expectations. Like in Muzaffarnagar [13], fake radical messages by some radical organizations are spread that causes serious violence in that area, many people are killed and forced to leave their houses. Such incidents are also occurring in Australia and in some other countries. Radical organizations play with peoples emotions and use it to make their political dreams come true.

2. Ideological Factors

Some radical organizations have some different ideologies about the other religions which create hated perceptions about these religions. These people want to prove that their ideology is best as compare to the others and tried to implement that forcefully over the others. In some cases, some political organizations are also spread such content over the internet to create hated perception among the people to fulfil their political expectations.

3. Human Psychology

In most of the terrorist organizations, hated religious content about their religion is spread among the specific religion peoples to recruit them for the violent activities. In that they completely brainwash the peoples and play with psychology and emotions to put their view into them and recruit them for the violent activities.

Like a hated content in which people were killed by the other religion's peoples are sent to the people which they want to recruit.

Some other factors are also there which going to influence the violent extremism over the internet.

Detection and prediction of such activities is a necessary task to do. That provides prior knowledge about such activities. That helps to make policies to restrict such violent extremist recruitment.

A brief literature review over the techniques which used to predict about such activities is presented in section II Literature Review. There are techniques like LDA [1] (Latent Dirichlet Allocation), CTM [6] (Correlated Topic Model) etc. are used to predict about violent extremist recruitment. Time series models Like ARIMA [1] are used to forecast such activities.

II. Literature Review

In [1], forecasting of violent extremist detection technique in cyber recruitment is presented. In that technique, LDA (Latent Dirichlet Allocation) is used to analyze the web content and predict web content. Then a forecasting model is used to provide better performance in that environment. And ARIMA (Auto Regressive Integrated Moving Average) based technique is used to provide better performance to the user. On that basis a forecasting mechanism for the violent extremist detection is provided to the user to detect extremist comments of the various recruitment processes.
But there is a problem with this technique, it is not able to form proper relations among the topics, which degrades the performance of the whole system to predict data.

In [2], a discussion about the violent extremism over online social media, which supports government to make plans to restrict such activities. In current scenario, there are various messages or other contents are uploaded over the social media or other online forums. These contents are spread rapidly across the border of the countries. That creates serious issues for the safety of the country. Thus a policy is required to restrict such rationalism over the internet, either by the Government or ISPs (Internet Service Providers). Safety measures like security and surveillance is required to restrict such activities.

In [3], a forum crawler which is used to analyze the content of the forums to detect extremist comments in the data is presented. There is an analysis over the iROBOT a forum crawler is presented. Which generally used to detect violent extremist and rational comments over the forums. And an enhanced technique is also presented which provides enhanced functionality to detect such violent extremism over the forums.

In [4], a data extraction technique from the various forums is presented. In online forums a huge amount of data is shared which requires better technique to extract useful information from that data. A lightweight and generalized algorithm is presented which provides better performance to extract data from the web forums. That data contains information about the rational comments and some other extremist type of data and can be used to recruit people from all over the world for violent extremist activity.
In [5], a method to detect terrorist related content over the online social media is presented. In that technique a machine learning technique is used to provide a better extraction mechanism for the data, is presented. There are two features called data dependant features and data independent features are used to extract data from the social media. Data dependent features, are the features which are influenced by the dataset. Data independent features are the features which can be used on the other dataset with similar results. In that, baseline approach and AdaBoost classifiers are used to provide better performance to extract data from these forums.
In [6], Correlated Topic Model technique is presented. In existing techniques, LDA based technique is used to analyse and predict data, but that technique is not able to provide realistic relation among the topics thus a CTM (correlated topic model) is presented, which provides better performance to the user to form relation among the various topics.

In [7], a technique to detect violent extremist recruitment in a social media communities is presented. In that a comparison of the techniques like Naïve Bayes, SVM (Support Vector Machine), Logistic Regression, Classification Trees etc. which are used for the purpose to detect violent extremist detection, is presented. In current scenario a huge amount of data is flooded over the social media sites. That data used to form cyber communities over the internet. These cyber communities attracts various radical organizations to recruit people for the violent extremist activity. To detect such activities there are several methods like supervised or unsupervised techniques are used. Data from the networks like AnsarAl-jihad network is used to detect violent extremists comment in the dataset. Thus an enhanced technique is provided to detect such extremist attacks.

III. Conclusion

A brief over the techniques used to detect violent extremism in social media is presented in this paper. Various machine learning techniques and Natural Language Processing (NLP) techniques like Correlated Topic Model (CTM), Latent Dirichlet Allocation (LDA) etc. are used to detect violent extremism content in social media posts which spread by the various radical organizations. ARIMA based forecasting technique also used to forecast about such recruitment. An enhanced technique which uses Doubly CTM and N- gram based model to detect violent extremism is proposed for the future work to enhance the performance of violent extremism detection in social media posts.

References

[1] Jacob R. Scanlon, Matthew S. Gerber,"Forecasting Violent Extremist Cyber Recruitment", IEEE, 2015.
[2] Murdoch Watney,"Intensifying State Surveillance of Electronic Communications: A Legal Solution in Addressing Extremism or Not?" IEEE, 2015.
[3] Hwei-Ming Ying, Vrizlynn L. L. Thing,"An Enhanced Intelligent Forum Crawler", IEEE, 2012.
[4] Wee-Yong Lim, Vyjayanthi Raja, Vrizlynn L. L. Thing "Generalized and Lightweight Algorithms for Automated Web Forum Content Extraction", IEEE, 2013.
[5] Lisa Kaati, Enghin Omer, Nico Prucha, Amendra Shrestha "Detecting Multipliers of Jihadism on Twitter" IEEE, 2015.
[6] David M. Blei, John D. Lafferty,"Correlated Topic Models", 2006.
[7] Jacob R Scanlon, Matthew S Gerber,"Automatic detection of cyber-recruitment by violent extremists", Security Informatics, 2014.
[8] [Online] Available: https://www.publicsafety.gc.ca/cnt/ntnl-scrt/cntr-trrrsm/cntrng-vlnt-xtrmsm/index-en.aspx.
[9] [Online] Available: http://www.state.gov/j/cve.
[10] [Online] Available: http://www.globalcenter.org/topics/countering-violent-extremism.
[11] [Online] Available: http://www.livingsafetogether.gov.au/aboutus/Pages/what-is-violent extremism.aspx.
[12] [Online] Available: http://www.gsdrc.org/professional-dev/violent-extremism/.
[13] [Online] Available: https://en.wikipedia.org/wiki/2013_Muzaffarnagar_riots.

Aditi Saraf Jain received Bachelor of Engineering Degree in Computer Science and Engineering from Oriental College of Technology, Bhopal, M.P. in 2014. She is currently pursuing M.Tech from School of Information technology, Rajiv Gandhi Proudyogiki Vishwavidhyalaya, Bhopal in Computer Technology and Applications. Her research interest includes Violent Extremist Detection in Social Media Websites.

Dr. Jitendra Agrawal was born in 1974, is an Assistant Professor in the Department of Computer Science & Engineering at the Rajiv Gandhi Proudyogiki Vishwavidyalaya, MP, India. He earned his Master Degree from Samrat Ashok Technology Institute, Vidisha (M.P.) in 1997 and awarded Doctor of Philosophy in Computer & Information Technology in 2012. He has published more than 50 publications in International Journals and Conferences. He has published two books named Data Structures and Advanced Database Management System. He is the recipient of the Best Professor in Information Technology award by the World Education Congress in 2013. He is a senior member of the IEEE (USA), Life member of Computer Society of India (CSI), Life member of Indian Society of Technical Education (ISTE).

Dr Shikha Agrawal is an Assistant Professor in Department of Computer Science & Engineering at University Institute of Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal (MP) India. She obtained B.E., M.Tech. and Ph.D in Computer Science & Engineering from Rajiv Gandhi Proudyogiki Vishwavidalaya Bhopal. She has more than twelve years of teaching experience. She has been awarded as "Young Sciencetist" by Madhya Pradesh Council of Science and Technology, Bhopal in 2012. Her other extraordinary achievements include "ICT Rising Star of the Year Award 2015" and "Young ICON Award 2015". Her area of interest is Artificial Intelligence, Soft Computing and Particle Swarm Optimization and Database. She has published more than 30 research papers in different reputed international journals and 9 chapters. She is a Madhya Pradesh State Students' Coordinator, Computer Society of India (CSI), 2016-2017.