



# AN OPTIMAL CTM BASE APPROACH FOR SOCIAL VIOLANCE ANALYSIS OVER LDA

Priyanka Ustrate, Prof. Damodar Tiwari

MTECH CSE, BIST BHOPAL

<sup>1</sup>truthfuldesires@gmail.com

## Abstract

*In current scenario social media and other online applications provides an easy and fast way to communicate to the others. But some radical organizations use these mediums to spread wrong perception among the society and to recruit peoples for terrorist and violent extremist activities .these organizations use these mediums to recruit youth to conduct their terrorist movements. A CTM based technique is presented which provides better detection for such posts which are used for online cyber recruitment by the radical organizations or terrorist groups. A result analysis for the proposed technique is presented in IV Result Analysis section which shows that CTM based technique provides better results as compare to the existing techniques.*

**Keywords:-** SVM (support vector machine), Voilent Extremist, White power music, LDA(Latent Dirichlet allocation), CTM (Correlated Topic Model).

## I Introduction

In now days, social media and social blogs and other such components of web provide an easy and powerful platform to share information and communicate with other peoples. But some radical and terrorist organizations uses these techniques to influence people and recruit youth, to conduct their activities. There are several virtual communities are presents on the web which used to conduct such activities. Like an online magazine published by al-Qaeda called Arabian Peninsula, such magazines or other web contents spread wrong perception among a certain community and used to influence people to join in their terrorist movements. There are many activities are conducted to for online cyber recruitment for terrorist activities.

There various techniques are presented to detect such activities like, in [4] a description of the language which is used for such recruitment like

Azan, is presented. In [5] effect of the counter terrorism policies over the youth, how youth join in such activities and role of policies to which generate wrong perception among the youth, is presented. [10] a review over the various techniques which are used to detect radicalization of the youth, is presented. In [11] a technique is presented which uses to identifies the intensions of posts in the forums, in [12] [13] [14] detection scheme to detect cyber terrorism activities are presented.

To detect such violent extremist recruitment activity a [1] LDA (Latent Dirichlet Allocation) based technique is used which uses ARIMA model to forecast such activity. For that a social media contents are used to analyze such activity. But LDA uses dirichlet distribution to map content to predict such activities which is not able to provide accurate result, because it not take relation in count to map topics. To reduce this defect a CTM (Correlated Topic Model) based technique is presented which used logistic normal distribution to map topics. That technique provides better and accurate performance to map topics. A result analysis for the propose technique is presented in result and analysis section, which shows that proposed technique provides better result as compare to the existing technique.

## II Related Work

In [1] a forecasting technique which used to forecast the activity of violent extremist recruitment in forum is presented. In that technique a SVM based model used to detect recruitment post in the forum. A LDA (latent dirichlet allocation) is used to analyze content of the post in the forum. Put that in to different time series model to forecast that recruitment process. That technique provides less no of forecasting error as compare to the other existing technique. In this paper an ARIMA (Auto regressive integrated moving average), PCR

(Principal component regression) is used in the existing technique but there is a problem in generation of the prediction of the recruitment. In this paper an ARIMA and ETS (Exponential smoothing) is used to provide the better forecasting results. Comparison with other technique shows that this technique provides better results as compare to the other technique.

In [2] a SVM (Support Vector Machine) based classification technique is presented which used to identify the post related to the violent extremist recruitment. To analyze that process a dark web portals data is used which contains data of more than 28 social websites. That data poses the content related to violent extremist activity and that data also contains religious (Islamic) discussion. In that an analysis process on the basis of the different factors likes the time frame, data sources and some other factors. SVM classification technique is used to classify that data which helps to identify the violent extremist recruitment activity. In previous technique a naïve based classification technique is used to classify that data which is not efficient to provide better results.

In [3] perspective over the virtual communities supporting terrorist and violent extremist activities is presented. These virtual communities play a vital role in such extremist recruitment processes. Virtual communities or group over social networks poses components like white power music and white supremacist games which used to promote the racism against the non-whites are used to conduct activities to recruit peoples for violent extremist activities. There are many other means by which radical perception is created, which used to influence the peoples for the terrorist activities.

### **III Proposed Work**

In existing techniques, LDA based prediction scheme is used to detect and forecast any violent recruitment activity in forum or online websites. For that first data of that post is classify using SVM (Support Vector Machine), then LDA (Latent Dirichlet Allocation) to prediction purpose and then the results of the LDA provided into ARIMA (Auto Regressive Integrated Moving Average) is used to provide a forecast of that data.

But in [9] a comparison analysis of the LDA and CTM (Correlated Topic Model) is shown which shows that CTM provides better results as compare to the LDA. In LDA a dirichlet distribution is used because of that the data of the topics varies, on the other hand in CTM based techniques a Logistic normal distribution is used to form the topics relation. In dirichlet distribution independent proportion among component is found thus there is no relation with the other topics is found. On the other hand in CTM a new Logistic normal distribution is used, which provide relation among the topics and provides a flexible framework for the process.

In proposed technique a CTM based technique is presented to forecast violent extremism in the online websites or in social media. In that first SVM is used to classify data then a CTM based technique is used to predict violent or radical recruitment. Then ARIMA model is used to forecast these activities. A comparison analysis for the proposed method is shown in IV Result Analysis section.

### **IV Result Analysis**

In this section a result analysis for the existing LDA based forecasting technique and proposed CTM based forecasting technique is presented. For analysis purpose forecasting values and actual values which generated by both the techniques are used as parameter.

In Table 4.1, a statistical comparison of the forecasted values which forecasts violent extremist online recruitment in social media or website's posts which are posted by radical or terrorist organizations. A graphical analysis is shown in Graph 4.1, which shows that CTM based technique provide better forecasting results as compare to existing LDA based technique.

In table 4.2, a statistical analysis of actual posts which are posted for violent extremist recruitment. That analysis shows, CTM based technique provides better detection results as compare to the existing LDA based technique.

**Table 4.1: Violent extremist statics for forecasting values**

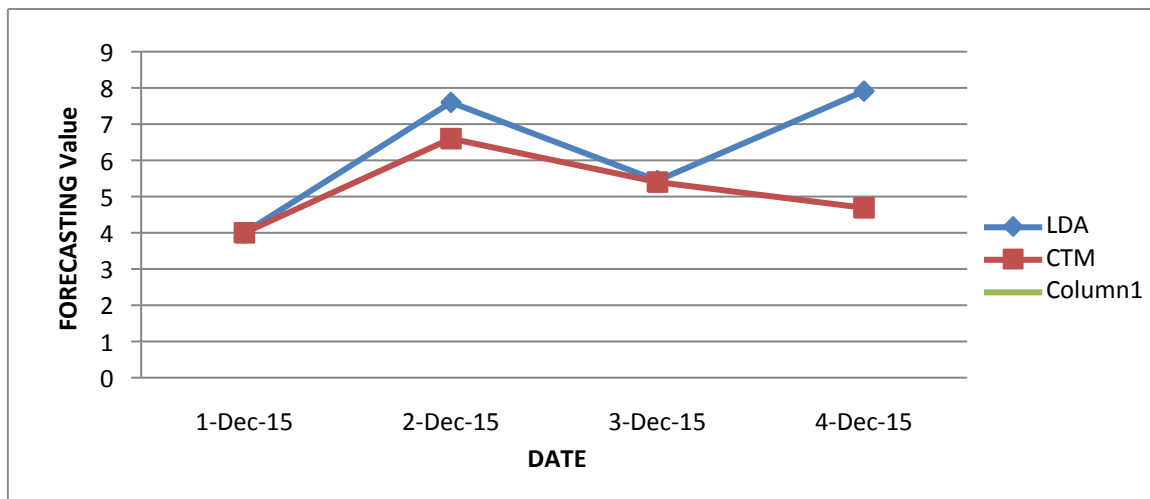
Date	LDA	CTM
01-12-2015	4	4
02-12-2015	7.60	6.60
03-12-2015	5.44	5.40
04-12-2015	7.91	5.91
05-12-2015	5.72	4.69

A graphical analysis is also presented in Graph 4.2, which shows that CTM based forecasting technique is efficient to provide better detection for the violent extremist recruitment detection in post of online content.

Forecasting values:- it is the value, generated by the use of ARIMA model, which predict about the attacks or any type of violent extremist activity.

Actual Value:- It is the, within a time span how many violent extremist activities are performed.

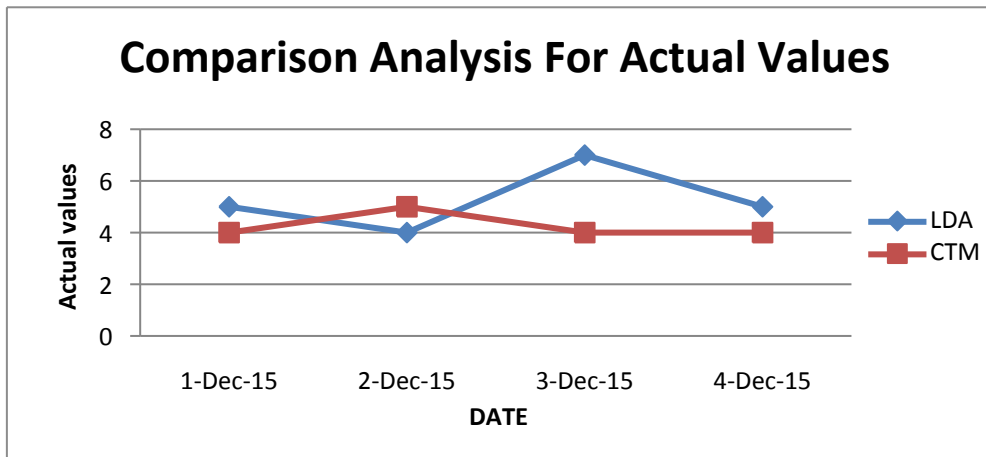
Evaluation parameters



Graph 4.1: Graphical analysis for the forecasting values.

Table 2: Violent extremist Statics for actual values.

Date	LDA	CTM
01-12- 2015	5	4
02-12-2015	4	5
03-12-2015	7	4
04-12-2015	5	4
05-12-2015	5	3



Graph 4.2: Graphical analysis for actual values.

Both graphical and statistical analysis shows that CTM based forecasting technique provides accurate result as compare to the LDA based technique. Propose technique is efficient to deal with such violent extremist recruitment activities.

#### V Conclusion

Due to huge evolution of the internet there are huge amount of data is posted over the internet. Some radical and terrorist organizations use these posts to recruit youth to conduct their activities. A CTM based technique is presented in this paper to predict about such posts to restrict such activities. In section IV, a comparison analysis with existing LDA based technique is presented which shows that proposed technique performs better as compare to the existing techniques.

#### VI References

- [1] Jacob R. Scanlon and Matthew S. Gerber "Forecasting Violent Extremist Cyber Recruitment" IEEE Vol. 10, No. 11, November 2015.
- [2] Jacob R Scanlon and Matthew S Gerber "Automatic detection of cyber-recruitment by violent extremists" Scanlon and Gerber Security Informatics 2014.
- [3] Lorraine Bowman-Grieve "A psychological perspective on virtual communities supporting terrorist & extremist ideologies as a tool for recruitment" Security Informatics 2013.
- [4] David B Skillicorn1 and Edna F Reid "Language use in the Jihadist magazines inspire and Azan" Security Informatics 2014.
- [5] Margarita Bizina, David H. Gray "Radicalization of Youth as a Growing Concern for Counter-Terrorism Policy" Global Security Studies, Winter 2014.
- [6] Rasmus Rosenqvist Petersen and Uffe Kock Wiil "CrimeFighter Investigator: Integrating synthesis and sense-making for criminal network investigation" Security Informatics 2013.
- [7] K. Peters and P. Richards, "Why we fight": Voices of youth combatants in Sierra Leone," J. Int. African Inst., vol. 68, no. 2, Apr. 1998.
- [8] J. M. Weinstein, Inside Rebellion: The Politics of Insurgent Violence. New York, NY, USA: Cambridge Univ. Press, 2007.
- [9] Sango Lee, Jaeki Song, Yongjin Kim "An Empirical Comparison Of Four Text Mining Methods" journal of Computer information systems, 17 February 2010.
- [10] DENZIL CORREA and ASHISH SUREKA "Solutions to Detect and Analyze Online Radicalization : A Survey" IITD, January 2013.
- [11] Z. Chen, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, "Identifying intention posts in discussion forums," in Proc. NAACL-HLT, Jun. 2013.
- [12] L. A. Overbey et al., "Virtual DNA: Investigating cyber-behaviors in virtual worlds," Space and Naval Warfare System Center Atlantic, Charleston, SC, USA, Tech. Rep. 33-09E, 2009.
- [13] G. S. McNeal, "Cyber embargo: Countering the Internet jihad," Case Western Reserve Univ. J. Int. Law, vol. 39, pp. 789–826, 2008.



- [14] H. Chen, S. Thoms, and T. Fu, “Cyber extremism in Web 2.0: An exploratory study of international Jihadist groups,” in Proc. IEEE Int. Conf. Intell. Secur. Inform. (ISI), June 2012.
- [15] M. Yang, M. Kiang, H. Chen, and Y. Li, “Artificial immune system for illicit content identification in social media,” J. Amer. Soc. Inf. Sci. Technol., vol. 63, no. 2, pp. 256–269, 2012.
- [16] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet allocation,” J. Mach. Learn. Res., vol. 3, pp. 993–1022, 2003.