



Survey Paper on Topic Modeling Techniques to Gain Useful Forecasting Information on Violent Extremist Activities over Cyber Space

Ashutosh Sharma, Dr. Mahesh Pawar

School of Information Technology, RGPV University,
Bhopal, India

Abstract- Recent advances in communications technology are providing a medium for individuals or groups to subscribe to extremist worldviews and form networks, access training and obtain information, while remaining virtually undetected in the online world. While the Internet is facilitating global virtual communities like Second Life, MySpace, Facebook and Twitter it is also providing an anonymous meeting place for dissatisfied individuals to gather, share ideas, post and exchange information regarding their particular ideology. This virtual community provides a sense of belonging to a global cause in which the actions of an individual can be aligned to, and seen to contribute towards something more significant than their own lives. Membership of this virtual community can facilitate the training of individuals, thereby negating psychological barriers that would normally inhibit particular types of behaviour. Terrorist groups operate as unstructured, fluid networks providing them significant advantages over rigidly structured state and nation based law enforcement agencies. In addition terrorist groups are exploiting the combination of rapidly evolving technology and legislation to prevent detection. This review could ultimately help identify the impact of violent organizations, like terrorist groups, within the social network of an online community.

Index Terms- Terrorism, Internet, Training, Radicalisation, Data Mining, Social Networking, Terrorist recruitment, Intelligence, Internet Archive, LDA(Latent Dirichlet allocation).

I. INTRODUCTION

In the last few years extremism has expanded to internet and online social media [1,2]. Extremist organization uses techniques to hire new terrorist through social media. According to recent research cyber tools are more influential at the start of a future member's extremist activity- the recruitment and radicalization phase.

Basically terrorist uses free and open nature of internet online communities [3] and they distribute literature and training materials [4,2]. Terrorist organizations spread awareness for recruiting members on social websites like Second Life, Facebook, twitter and radicalized religious web forums [5,1,2]. Through this intelligent community would benefit for knowledge whom to target and how terrorist conduct online recruitment.

FBI investigation report cited a "data explosion" and "workload" as contributing factors to analyst and agent oversights[6]. At "nearly 20,000 Aulahi-related [electronics documents], "keeping up with workload demands was clearly a challenge for the two reviewers assigned to the case at the time[6]. This large amount of data require review which is done by limited no of personnel, automated methods can be of use for pre-screening text documents and reducing the workload of human analysts. Automated forecasting method can be of use for anticipating future workload of human analyst and prescanning text document.

Violent extremists of all have become increasingly proficient using the internet and social media to propagate their ideologies and radicalize and recruit a generation that has grown up online. In this survey, a violent extremist (VE) uses violent means to disrupt legitimate authority and spread terrorism. Insurgents and terrorists violent extremist are the organisation that spread ideology of hatred and instigate violence. A radical group organizing a peaceful protest are not considered as terrorist. Many modern groups, like the Westboro Baptist Church, have radical religious views, but these beliefs are neither necessary nor sufficient to classify them as violent extremists without the intent to carry out or advocate for specific acts of violence. In this survey, VE recruitment is any attempt by a group or individual involved in VE to recruit, radicalize, or persuade another person to aid a violent extremist movement. VE cyber-recruitment is therefore VE recruitment activity that makes use of computers and the internet easiest means to recruit terrorist and spread hatred among the society by messenger.

II. LITERATURE REVIEW

(S. Gerber et. al.)[10] The use of internet in past few years has increased exponentially and hence made it a platform for millions therefore if any one wants to reach millions at once then internet is their prime target which resulted in formation of several communities over internet. This has attracted violent extremist groups towards cyber space to conduct their unlawful acts or to prepare for them.

R. Torok et. al. [2] This paper suggests how al-Qaeda attempts to use cyber technologies to recruit and engage home grown terrorists. Deliberately, al-Qaeda is shaping an endless arrangement of web systems to give a stage to achieve an

overall group of onlookers in part, of using internet technology, terrorist groups have become more network oriented in their structures

social network Informal community examination shows that comparative individuals tend to locate one another . Al-Qaeda and different gatherings attempt to make that procedure as simple as possible

Durgesh K. Srivastav et. al. [25] In this paper, a novel learning system, Support Vector Machine (SVM), is connected on diverse information which have two or multi class. SVM, a capable machine technique created from understanding the statistical learning which undoubtedly has made critical accomplishment in some field SVM method does not suffer the limitations of data dimensionality and limited samples Several studies have concluded that the SVM (support vector machines) are more efficient and are better in terms of overall performance when compared in terms of classification accuracy then other data mining and analysis algorithms for classification of data .

Deerwester S. et. al. [16] In this paper, the author explained a new approach to gain information of implicit higher-order structure to relate words with documents that is their semantic structure in order to improve the detection of related documents in context of words found in queries. The latent semantic indexing (LSI) analysis that they have tried uses singular-value decomposition. They took a large matrix of word and document relationship data and construct a “semantic” space wherein terms and documents that are closely associated are placed near one another. Singular-value decomposition allows the arrangement of the space to reflect the major associative patterns in the data, and ignore the smaller, less important influences. As a result, terms that did not actually appear in a document may still end up close to the document, if that is consistent with the major patterns of association in the data. Position in the space then serves as the new kind of semantic indexing. Retrieval proceeds by using the terms in a query to identify a point in the space, and documents in its neighbourhood are returned to the user.

Thomas Hofmann et. al.[20] This paper describes the use of PLSA (Probabilistic Latent Semantic Analysis) PLSA is closely associated with LSA(Latent Semantic Analysis) it is a statistical method for factor analysis of binary and count data. In contrast to LSA which uses linear algebra and performs a single value decomposition PLSA uses a generative latent class model to perform a probabilistic mixture decomposition. Probabilistic Latent Semantic Analysis has many applications, most prominently in information retrieval, natural language processing, machine learning from text.

This paper concluded that PLSA is more decent approach then the existing LDA as it has a detailed statistical foundation and uses a likelihood function as optimisation criteria. It gained a decent performance gain in all cases.

David M. Blei, et. al. [16]This paper explain the problem of modeling text corpora and other collections of discrete data. The goal is to find short descriptions of the members of a collection that enable efficient processing of large collections while preserving the essential statistical relationships that are useful for basic tasks such as classification, novelty detection, summarization, and similarity and relevance judgments. For this LDA is proposed which is a generative probabilistic model for accumulations of discrete information, for example, content corpora. LDA is a three-level progressive Bayesian model, in which every thing of an accumulation is demonstrated as a limited blend over a hidden arrangement of subjects. Every subject is, thusly, displayed as a boundless blend over a basic arrangement of theme probabilities. In the setting of content displaying, the theme probabilities give an express representation of an archive.

H. Chen,et. al. [7]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.

S. Thoms et. al.[8] 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.

Lorraine Bowman-Grieve et. al.[12] 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.

David B Skillicorn1 and Edna F Reid [13] A study over the language used in the Jihadists magazines and other forums, and their influence and impact over the society is presented. This paper focuses three aspects one is how high level descriptors are used in the language to influence the peoples, second computational methodology to validate effectiveness of the language, third how these jihadist magazines are going to influence the society. In this a language model is used to analyze the high level property of the document and provide the data or words in a corpus.

Process to analyze that data like occurrence of such influential words and treated as parameter to judge any document.

David H. Gray et. al.[14] Focuses on the problem of radicalization of the youth in the western countries. In that most of the youth are second generation muslim born in western countries thus they not required any visa to travel this country. In that case more and more people born in countries like USA, UK, and Canada are going involve in many terrorist activities. By conduct many counter terrorism activities in these countries generate insecurity in the youth of muslim community and that create a perception of negligence and by the speeches of the some influential leaders and radical organizations they fall to such activities.

Rasmus Rosenqvist Petersen et. al.[15] A human centered and target centric model is presented that provides an efficient way for the crime investigations. In this costumer generates a request to get information about any target.

For further process Investigator request for the information of the target to the collectors. Then the information related to the target is provided by the collectors but that information in the form of multiple pieces then investigator arranges that data pieces and create a model for the investigation. Then information related to the target extracted from that. In that way a crime investigation process is conducted.

III. COMPARISON TABLE OF TECHNIQUES USED.

Models	Characteristics	Limitations
Latent Semantic Analysis (LSA)[16,17]	<ul style="list-style-type: none"> Reduces dimensionality of tf-idf using Singular Value Decomposition. Captures synonyms of words. Not robust statistical background. 	<ul style="list-style-type: none"> Difficult to determine the number of topics. Difficult to interpret loading values with probability meaning. Difficult to label a topic in some cases using words in the topic
Probabilistic latent semantic analysis(PLSA) [18,19,20,21]	<ul style="list-style-type: none"> Mixture components are multinomial random variables that can be viewed as representations of “topics.” Each word is generated from a single topic; different words in a document may be generated from different topics. PLSA partially handles polysemy. 	<ul style="list-style-type: none"> No probabilistic model at the level of documents
Latent Dirichelet Allocation (LDA) [22,23]	<ul style="list-style-type: none"> Provides full generative model with multinomial distribution for words is topics and Dirichelet distribution over topics. Handles long-length documents shows adjectives and nouns in topics. 	<ul style="list-style-type: none"> Incapable to model relations among topics
Correlated Topic Model (CTM) [24]	<ul style="list-style-type: none"> Considers relations among topics using logistic normal distribution Allows the occurrences of words in other topics Allows topic graphs 	<ul style="list-style-type: none"> Requires complex computations Contains too general words in topics.

IV. STRATEGIES OF FORECASTING VIOLENT EXTREMIST

4.1. Current Insights into Violent Extremism:

Marc Sageman’s paper emphatically states that extremism is not the problem, violent extremism is the problem. He describes the trajectory of Islamist violent extremism from a protest community/social movement based on political grievance and neo-jihadi ideology, which frames injustice related grievances using the “West is at war with Islam” narrative, to a much smaller group of individuals who act out violently, motivated by moral outrage, small group dynamics, and norms. He counsels that a multi-layered approach is required to deal with violent extremism that avoids repression (which leads to violence), addresses discrimination, attacks the disconnects between the fantasy of being a violent extremist and the reality, and leverages the controversy surrounding issues like violence against civilians. Sageman also reinforces the need for consistency of action with words and the fair treatment of Muslims. He highlights the potential for using the Internet for effective counter messaging by credible individuals (something expounded on in the paper by Speckhard in Section 4).

When messaging, it is critical to understand the narratives in play and how they are manifested in local, national, and global discourse. Steven Corman writes of the importance of narratives, which are systems of stories that house themes, forms, and archetypes. Based on a study of Islamist extremist narrative, thirteen “master narratives” were identified. For example, the “nakba” or “catastrophe” narrative represents the stories of the loss of Palestine and Jerusalem as well as betrayal and injustice themes, implicitly calling for deliverance. In another example, the “Crusader” narrative describes the ultimate victory over an occupier, implicitly calling for a champion. Master narratives are those that are frequently retold over time. Corman relates that master narratives are powerful because they effectively form a bridge between a story and a present day event/situation. He argues that, because narratives provide an alternative form of rationality or cognitive lens, effective counter messaging can be developed which challenges narratives based on questioning their narrative coherence

or structural consistency (plausibility of story, consistency of event sequences) or narrative fidelity or truth (plausibility based on the reader's experience). He states that narratives that draw on master narratives are a special case in terms of counter messaging, requiring discrediting the argument that links a present day situation and the narrative stories by pointing out differences or by providing an alternative interpretation

4.2.Prevention of Violent Extremism:

Much of the current effort in countering violent extremism is focused on security and law enforcement. However, research has shown that merely removing violent extremists (whether by killing, imprisonment, or relocation) is not effective as the numbers of new recruits will dwarf the numbers removed, and civilian casualties from security/military operations have similar effects on recruitment and radicalization.⁶ The goal of preventing violent extremism is to eliminate or minimize those factors that lead individuals to join violent extremist organizations or to support violent extremism. The contributions in this section offer a variety of perspectives, based on research, as well as firsthand experience with violent extremism, on the prevention of violent extremism. These contributions include a method to determine appropriate populations to target for prevention strategies; a strategic plan for systematically addressing the Islamization process and components; suggestions of trends to exploit and perceptions/mistakes to avoid in order to prevent violent extremism; the importance of and issues associated with balancing security solutions (which treat symptoms) with a variety of development solutions that focus on the causes of extremism; and the importance of cultivating appropriate partners and supporting a variety of solutions that engage, empower, and foster the ownership of Muslim communities and leaders

4.3.Delegitimization and Other Strategies for Minimizing Support for Violent Extremism:

The vast majority of violent extremist organizations disappear within a few years of their formation.⁹ The ones that survive are typically those that are able to secure funding/support from various sources.¹⁰ Thus, any set of strategies to counter violent extremism must include strategies for countering the supporters of violent extremist organizations. Martha Crenshaw writes that for "terrorism to appear legitimate, there must be a congruence between ends and means; both the resort to terrorism and the particular form it takes should seem appropriate to the cause."¹¹ Due to a vacuum left by a dysfunctional or ineffective government, violent extremist organization may, themselves, aspire to legitimacy by providing good governance (e.g., the Taliban provides parallel governance including services and rule of law). One such strategy is to delegitimize the violent extremist organization and their actions in order to make the organization less credible and less likely to provide alternative solutions for existing grievances. Delegitimization and other solutions, including use of popular culture (e.g., television or music) to provide countervailing images and messages, are discussed in this section.

4.4.Counter Violent Extremism Strategies for AffectingBehaviorChange:Deradicalization/ Disengagement, Mediation, and Influence:

It is not possible to prevent every individual or group from turning to violent extremism. Thus, it is necessary to have strategies that seek to change the violent behavior (i.e., motivate the disengagement of the individual or group from violence against a particular entity or violence in general) or change the belief (deradicalization) such that an individual is no longer an extremist or at least not a violent one. The bulk of the research on violent extremism has focused on recruitment and radicalization rather than deradicalization and disengagement. Moreover, although other literature has focused on disengagement from cults, gangs, etc., there is still much to be learned about why violent extremists leave the group and/or abandon violence. One thing we do know, however, is that most of those who are involved in terrorism (or violent extremism) disengage in various ways.¹² This section motivates the need for a better understanding of how violent extremism ends, a more critical evaluation of existing deradicalization/disengagement methods, and discusses several

V. DETERRENCE AND INFLUENCE

The traditional meaning of deterrence refers to successfully avoiding an action by another party by threatening to punish that party if the action is taken (and, implicitly, foregoing that punishment if the action is not taken). In its most familiar form, the superpowers during the Cold War achieved a state of mutual deterrence by having the assured and credible capability to respond effectively to a nuclear attack by the other. People still disagree about what was necessary for deterrence, but the *countervailing strategy* that emerged after study in the Nixon, Ford, and Carter administrations was, arguably, the mature U.S. Cold War concept (Slocombe, 2003; Brown, 1983).

A classic deterrent strategy is unlikely to succeed against people such as Osama bin Laden or Ayman al Zawahari. First, the United States and their allies are already attempting to destroy the al Qaeda organization and to capture or kill them personally (it has now succeeded with bin Laden); it is unlikely that they would stop if the al Qaeda leadership merely promised a measure of future restraint. Deterring top leaders, then, is probably a non-starter. What, then, would make more sense?

As mentioned in the introduction, the answer is that it is far better to think in terms of a spectrum of *influences* (Davis & Jenkins, 2002). Classic deterrence is one such influence, but many others exist as well, as noted also by the late Alexander George (George & Smoke, 1973; George, 2003) in his work on coercive diplomacy. When referring to a generic VEO, the ways to influence include: (1) invite, (2) induce, (3) encourage, (4) discourage, (5) deter (by threat of punishment), (6) head off (by raising perceived risks and uncertainties), (7) be seen as able to defeat an attack (sometimes called deter by denial), (8) deter an additional attack by punishing now (perhaps to illustrate what punishment means or to demonstrate credibility), (9) deter the next attack by defeating now, or (10) deter the next attack by essentially crushing the other party

now. The first four of these do not involve the use of force or even threats regarding the use of force as others do. In any case, a wide range of options exists.

5.1 Seeing the VEO as a System

It is only natural to think of a VEO as a single entity. However, the al Qaeda system and other VEO systems include lieutenants (with a range of motivations), foot soldiers, logisticians, facilitators, recruiters, spiritual Leaders, and other theoreticians; they also include external supporters, whether nations, organizations within a nation (such as elements within Pakistan’s Inter-Services Intelligence), or individuals. And, importantly, the VEO “system” includes those in the public who support the VEO, either directly or passively (by turning a blind eye and tolerating their presence) (Davis, 2010). No one doubts that the effectiveness of a VEO depends on all of these components to greater or lesser degree, i.e., that such a system view is essential. All this means, however, that we should ask how a component could be deterred or otherwise influenced .

This observation about the need to decompose “the system” and consider which influences might be effective against the individual components has become more obviously crucial as the “al Qaeda” threat has morphed into something much more complex—al Qaeda Central and its many al Qaeda “affiliates.” The affiliates have greatly varied motivations, allegiances, and vulnerabilities. Further, the system is networked, which creates additional strengths, weaknesses, and targets for attention (Arquilla & Borer, 2007).

5.2. Insurgency, Revolution, and Terrorism

The United States has been embroiled in Iraq and Afghanistan for nearly a decade as part of what was earlier called the “Global War on Terrorism.” Counterinsurgency (COIN) has become a primary element of U.S. actions. It is, therefore, easy to confuse counterinsurgency with counterterrorism. Terrorism, however, is a tactic (and sometimes a strategy) that may or may not be used by insurgents (Hoffman, 2006).

To be concrete, for insurgents to plant improvised explosive devices (IEDs) to disrupt or destroy U.S. military convoys is extremely troublesome to U.S. counterinsurgent efforts, but it is not terrorism. Nor is it terrorism for combatants to kill enemy leaders, whether by sniper fire or by drone attacks. Nor is it terrorism to mount attacks on the other sides’ military forces with some accidental “collateral damage” in the form of civilian deaths or destruction of infrastructure. Such side effects are tragic and the United States goes to extraordinary lengths to avoid them, but such attacks are not terrorism (Massive firebombing of cities, as in World War II, was another matter.).

Making the distinction is important because the plausibility of influence is very different depending on what behavior one is attempting to prevent.

5.3. Factors Affecting Insurgency and Terrorism

A useful step in considering strategies to counter terrorism is understanding the factors that contribute to it. A recent study by Davis & Cragin, 2009 reviewed the social-science literature for insights about factors underlying terrorism; much of the resulting book also applies to insurgency, but the focus here is on terrorism. Figure 1 is one summary depiction from that work, expressed as a “factor tree.” Forthcoming work (Davis, Larson, et al., 2011), peer reviewed but currently undergoing security review for public release, has a richer version of this “factor tree” informed by further empirical study, but Figure 1 will suffice to illustrate the concept. The left branch of Figure 1 highlights motivations as an important factor in insurgencies and terrorism; often, terrorists and insurgents see themselves as part of a greater movement for a righteous cause. This cause may be a revolution inspired by repression and despotism, or rooted in extremist Salafi religious ideology, something noble such as defending one’s country against foreign invaders, or something very different, such as the excitement and glory of being involved in a violent group.

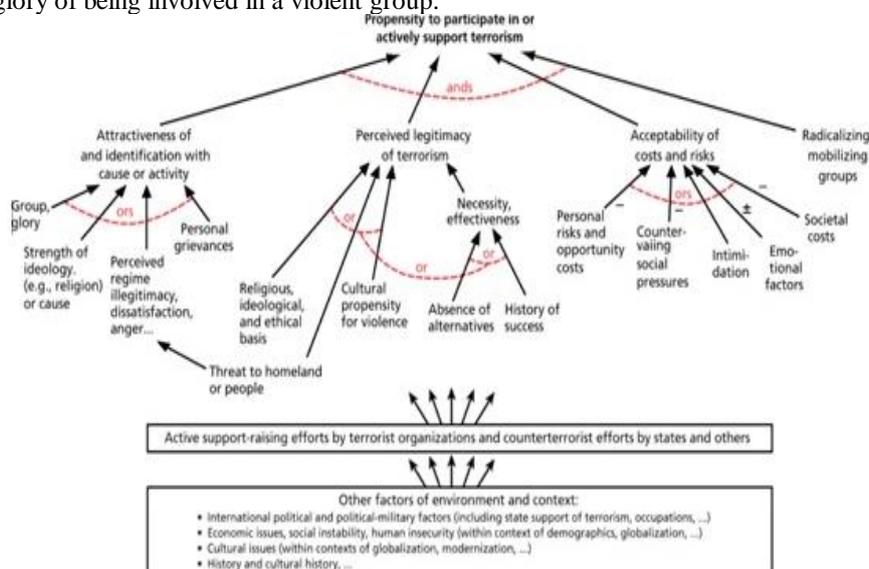


Figure 1 Factors affecting terrorism[26]

The figure also suggests that to participate in or support terrorism per se, not just joining a cause, also requires some sense of the legitimacy of terrorist actions. The basis for that sense may come from any of several sources, such as a religion or other ideology, a dire threat to the homeland, or necessity (the absence of other alternatives). In some cases, to be sure, terrorism may not need much rationalization because violence and brutality are seen by the relevant subculture as normal. Moving rightward in the figure, those participating or supporting terrorism are effectively making the judgment that doing so is worth the costs and risks. This “decision” may be neither conscious, explicit, nor “rational” in that people can be caught up in the emotions of revolution, for example, or too frightened to continue. People may also be greatly influenced by family, local respected personages, or icons. Finally (rightmost branch), a key factor is the existence of a mechanism, such as a mobilizing organization to join that has the leadership, resources, planning capability, and logistics actually to “do something.” Most societies have young “hot heads” willing to engage in violent acts; fortunately, they usually lack the mechanism for being effective.

VI. INFLUENCING THE COMPONENTS

Mobilizing Groups:

This factor tree can be used to think systematically about how to influence different aspects of a violent extremist system. Starting at the right, with a factor that is different in kind than the others, it is only reasonable to attack the “mobilizing mechanisms.” This could mean destroying infrastructure, killing or imprisoning leaders, repressing certain political parties, etc. The actions taken might be justified by law and carefully restrained, or might be the actions of a despot cracking down on a dangerous faction. This aspect of an overall deterrence strategy is better dealt with elsewhere and has little to do with influence—except that highly disruptive actions against an organizing mechanism can help dissuade or deter individuals from joining (they may deem it “not in their interest” or too dangerous). Such effects would manifest themselves in Figure 1 by reducing the perceived acceptability of costs and risks.

Motivations:

Moving now to the issue of motivations (left branch of Figure 1), it is straightforward to identify potential issues and contemplate how to influence matters positively. If the motivation for insurgency/revolution is to overthrow or change an inept and despotic government repressing the people, then governance needs to improve. This may or may not be something that the United States can affect. If a primary motivation is religious or otherwise ideological, then strategic communications to counter the extremist philosophy can play a role. In this connection, two cautions have strong support in the literature (Egner, 2009; Egner, 2010), as discussed also in other papers within the current volume.

- Strategic communications perceived to be from outsiders often have little credibility and can be counterproductive (exceptions exist, such as President Obama’s Cairo speech)
- Strategic communications need to be credible and persuasive, rather than blatant propaganda at odds with reality.

One seldom-tapped social science research topic relevant to influencing motivations is study of big-city violence by gangs. Such studies (e.g., Kennedy, Braga, & Piehl, 2001) are quite relevant to the “group and glory” sub-branch of Figure 1. They note that in multiple big cities, the number of murders has been greatly reduced by directly engaging the relevant groups, pointing out certain realities (such as that their “brothers” routinely snitch on other brothers and pick up with girl friends of those in prison), and asking questions such as “Just who is it that says it’s all right to kill innocent people in drive-by shootings? Not your mother or your grandmother.” Thus there is a moral component amidst others. A key element, however, is social pressure, such as the threat that, if a killing occurs and seems to be from a gang, then all the gang members will be hassled (legally, of course, as in punishing parole violations). From a theoretical perspective, this is a kind of collective punishment.

Legitimacy:

Addressing the factor of “perceived legitimacy” is similarly challenging and is usually best done by an individual or group from within the society. Interestingly, field researchers sometimes report that the concept of terrorism as being wrong and immoral does not resonate with those they interview. Instead, related questions are turned on their heads by interviewees, who mention state terrorism and the killing of innocents by U.S. air strikes. We should not expect people to agree with outsiders trying to make the argument that terrorism is bad. At the same time, they may come to accept the argument implicitly and change behaviors accordingly, as in withdrawing their support of terrorist organizations. Sometimes, the explicit reasoning is less than ideal, as when Muslim populations turn against al Qaeda because of its indiscriminate killing of Muslims labeled by al Qaeda as “apostates.” Nonetheless, there is reason to believe that rejection of terrorism (especially mass killing of innocent civilians) is potentially a universal value. After all, it is accepted across much of the globe, including in societies that were once quite brutal. However, as with so much in this domain, the positive influences will have to emerge from within the populations—perhaps with indirect assistance that does not compromise those with the positive messages.

Acceptability of Cost and Benefits:

The third top-level factor of Figure 1 is named acceptability of costs and benefits, rather than, for example, “cost-benefit calculations,” because both decisions and behaviors are often not based on calm, objective, cost-benefit assessments. In my view, the origins of the myth of rational-actor decision making being generally descriptive are three: (1) sloppiness in

defining “rational decision,” (2) the fact that people and organizations’ behaviors can, in fact, often be “understood” (or at least rationalized) by applying a kind of rational-actor model, and (3) the strawman argument that very few leaders and groups are irrational in the sense of acting randomly. A better term for describing actual decision-making is “limited rationality” (Davis, Kulick, & Egner, 2005). The research on this issue is extensive. The late Nobel Laureate Herbert Simon upset the economics profession a half-century ago with work on bounded rationality (Simon, 1982), pointing out that businesses could not, in reality, apply the prescriptions of decision theory because they lacked information and the necessary calculation capability.

VII. CONCLUSION

In this paper a survey over the different technique which uses to predict violent extremism in social media is presented. There are techniques like SVM classification and LDA (Latent Dirichlet Allocation) is used to present a model to predict violent extremism in the social media content. Dark web portal data which contains data about the racial content over the internet or any activity that create any violent movement and or terrorist activity.

A survey over the topic modelling algorithms is done in this paper and characteristics and limitations are analysed to be suited with the concern of gathering information that can be usefull in gaining knowledge about violent extremist activities in cyber space .and hence using the best combination of algorithm to get most meaningful digging of data through which a workable dataset can be created is required . SVM + LDA is been used with diffrent time series algorithm [11] .LDA lags in considering relationships between topics which are been modelled making the understanding of topic structure unclear. Corelated Topic Model defines topic relationships CTM cosiders that if two topics are not related then their content that is words also are not same in those two topics thus defining the relationship amongst two topics .

REFERENCES

- [1] L. A. Overbey, G. McKoy, J. Gordon, and S. McKittrick, Automated sensing and social network analysis in virtual worlds, in *Intelligence and Security Informatics (ISI)*, 2010, pp. 179-184.
- [2] R. Torok, \Make A Bomb In Your Mums Kitchen': Cyber Recruiting And Socialisation of `White Moors' and Home Grown Jihadists, in *Australian Counter Terrorism Conference*, November 2010, pp. 54-61.
- [3] M. Rogers, The psychology of cyber-terrorism, *Terrorists, Victims and Society: Psy-chological Perspectives on Terrorism and its Consequences*, (2003), pp. 77-92.
- [4] S. O'Rourke, Virtual radicalisation: Challenges for police, (2007).
- [5] S. Mandal and E.-P. Lim, Second life: Limits of creativity or cyber threat?, in *IEEE Conference on Technologies for Homeland Security*, 2008.
- [6] W. H. Webster, D. E. Winter, J. Adrian L. Steel, W. M. Baker, R. J. Bruemmer, and K. L. Wainstein, Final report of the William H. Webster Commission on the Federal Bureau of Investigation, counterterrorism intelligence, and the events at Fort Hood, Texas on November 5, 2009, tech. report, Federal Bureau of Investigation, 2012.
- [7] H. Chen, W. Chung, J. Qin, E. Reid, M. Sageman, and G. Weimann, Uncoveringthe dark web: A case study of jihad on the web, *Journal of the American Society for Information Science and Technology*, 59 (2008), pp. 1347{1359.
- [8] H. Chen, S. Thoms, and T. Fu, Cyber extremism in web 2.0: An exploratory study of international jihadist groups, in *IEEE International Conference onIntelligence Security Informatics (ISI)*, 2008, pp. 98{103.
- [9] T. Fu, A. Abbasi, and H. Chen, A focused crawler for dark web forums, *Journal of the American Society for Information Science and Technology*, 61 (2010), pp. 1213{1231.
- [10] Jacob R. Scanlon and Matthew S. Gerber “Forecasting Violent Extremist Cyber Recruitment” *IEEE* Vol. 10, No. 11, November 2015.
- [11] Jacob R Scanlon and Matthew S Gerber “Automatic detection of cyber-recruitment by violent extremists”*Scanlon and Gerber Security Informatics* 2014.
- [12] Lorraine Bowman-Grieve “A psychological perspective on virtual communities supporting terrorist & extremist ideologies as a tool for recruitment” *Security Informatics* 2013.
- [13] David B Skillicorn1 and Edna F Reid “Language use in the Jihadist magazines inspire and Azan” *Security Informatics* 2014.
- [14] Margarita Bizina, David H. Gray “Radicalization of Youth as a Growing Concern for Counter-Terrorism Policy”*Global Security Studies*, Winter 2014.
- [15] Rasmus Rosenqvist Petersen and Uffe Kock Wiil “CrimeFighter Investigator: Integrating synthesis and sense-making for criminal network investigation” *Security Informatics* 2013
- [16] Deerwester, S., Dumais, S., Furnas, G., Landauer, T., and Harshman, R., “Indexing by Latent Semantic Analysis”, *Journal of the American Society for Information Science*, 41 (6), 1990, 391-407.
- [17] Landauer, T.K., Foltz, P.W., and Laham, D., “An introduction to latent semantic analysis”, *Discourse processes*, 25, 1998, 259-284.
- [18] Ding, C.H.Q., “A probabilistic model for Latent Semantic Indexing: Research Articles”, *Journal of the American Society for Information Science and Technology*, 56 (6), 2005, 597-608.
- [19] Fuhr, N., “Probabilistic models in information retrieval”, *The Computer Journal*, 35 (3), 1992, 243-255.
- [20] Hofmann, T., “Probabilistic latent semantic indexing”, *SIGIR-99*, ACM New York, NY, USA, 1999, 50-57.

- [21] Papadimitriou, C.H., Raghavan, P., Tamaki, H., and Vempala, S., “Latent semantic indexing: A probabilistic analysis”, *Journal of Computer and System Sciences*, 61, 2000, 217235.
- [22] Girolami, M., and Kaban, A., “On an equivalence between PLSI and LDA”, *ACM New York, NY, USA*, 2003, 433 434.
- [23] Blei, D.M., Ng, A.Y., and Jordan, M.I., “Latent Dirichlet Allocation”, *Journal of Machine Learning Research*, 3, 2003, 993-1022.
- [24] Blei, D.M., and Lafferty, J.D., “A Correlated Topic Model Fall 2010 *Journal of Computer Information Systems 9 of Science*”, *The Annals of Applied Statistics*, 1 (1), 2007, 17-35
- [25] Data Classification Using Support Vector Machine Idurgesh K. Srivastava, 2lekha Bhambhu
- [26] Understanding and Influencing Public Support for Insurgency and Terrorism : Davis, Paul K ; Larson, Eric V ; Haldeman, Zachary ; Oguz, Mustafa ; Rana, Yashodhara