

**INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH
TECHNOLOGY****DISCOVER TROLLING IDENTITY IN SOCIAL MEDIA : REVIEW****Tayade Pooja Mahendra *, Shaikh Shafi Shadulla***M.Tech, Department of Computer Science and Information Technology, Dr. BAMU, Aurangabad
M.Tech, Department of Computer Science and Information Technology, Dr. BAMU, Aurangabad

DOI: 10.5281/zenodo.573510

ABSTRACT

Online social networks bring valuable information to millions of users-but their accuracy is based on the integrity of their user base. User contributions in the form of posts, comments, and votes are essential to the success of online communities. An Internet troll is a member of an online social community who deliberately tries to disrupt, attack, offend or generally cause trouble within the community by posting certain comments, photos, videos, GIFs or some other form of online content. You can find trolls all over the Internet-on message boards, in YouTube video comments, on Facebook, on twitter, in blog comments and everywhere else that has an open area where people can freely post to express their thoughts and opinions. Troll is quite damaging to one's online reputation. The insult troll is hater, plain and simple. And they don't even really have to have a reason to hate or insult someone. These types of trolls will calling them names, accusing them of certain things, doing anything they can to get a negative emotional response from them -- just because they can. This type of trolling can become so serious that it can lead to or be considered a severe form of cyberbullying.

KEYWORDS: Trolls, Negative Sentiments, SVM.**INTRODUCTION**

The content of a troll posting usually falls into one of several different categories. It may consist of an evidently idiotic contradiction of common knowledge, a deliberately offensive insult to the readers of a newsgroup or mailing list, or a broad request for trivial follow-up postings. The result of such postings is frequently a deluge of indignant responses. Sometimes, the follow-up messages posted in response to a troll can represent a large fragment of the contents of a newsgroup or mailing list for as long as several days or weeks. These messages are dispatched around the world to thousands of computers, dissipate network resources and costing money for people who pay to download email or receive Usenet news from internet. Troll threads also disappoint people who are trying to carry on substantive conversation. People post such messages to get awareness, to buckle conversation, and to make anxiety. The best response to a troll is **no** response forever who posted on everywhere. If you post a follow-up message, you are participating to the resulting uproar and most likely captivate the troll.

Every internet troll has a different backstory, and therefore different reasons for feeling the need to troll a community on the internet. They most of time feels depressed, angry, sad, jealous or some other emotion they may not be responsive of that's influencing their online behavior. What makes trolling so easy is that anyone can do it, and it can be done from a safe, inaccessible place as against to interacting with others in person. Trolls can leather behind their polished computers, go out trolling for unease, and after they're all done, they can carry on with their real lives without facing any real consequences. Trolling makes a lot of spiritless people feel stronger.

A troll, is someone who creates quarrel on sites like Facebook, Twitter and Reddit by posting messages that are particularly disputable or provoking with the sole intent of inflammatory an emotional response from other users. These type of messages are habitually distracting and take focus away from the subject at hand, sending a judicious communication down a rabbit hole of rudeness, personal attacks and jokes about your parents.



Online social media and networks have emerged as the principal forum for the public discourse. However, this open global forum is threatened by users that actively try to undermine its operation. Such users engage in discussions without the intention of constructively contributing to the dialogue, but rather to disrupt it, and they are commonly referred to as trolls. Social media is becoming a platform for people to reinvent themselves, allowing users to present themselves in the best manner possible to increase their 'perceived attractiveness'.

Facebook, Twitter, Reddit, Instagram, Tumblr or practically any social networking sides : Now that almost anyone can comment on a status update, reply to a tweet, converse in a community thread or send an anonymous question, trolling is absolutely everywhere that people can use to interact. Instagram is especially bad, because it's a very public platform that people use to post photos of themselves—inviting everyone and anyone to judge their appearances in the comment section.

Trolling relate a range of antisocial online behaviors that aim at disrupting the normal operation of online social networks and media. resist trolling is an important problem in the online world. Existing approaches depend on human-based or automatic mechanisms for identifying trolls and troll posts. The goal is to identify the targets of the trolls, so as to intercept trolling before it occurs. Trolls are hostile users who post or spread misleading, offensive or senseless information on the network.

Trolling and other negative behaviour on websites is widespread, ranging from subtly provocative behaviour to outright abuse. Publishers have sought to develop lively online communities, with high levels of user-generated content. They study suspended users and their affect on analyses performed on the social structure of a particular social media platform, Twitter and Facebook. A troll can disrupt the discussion on a newsgroup, disseminate bad advice, and damage the feeling of trust in the newsgroup community. Trolls can cause significant legal or illegal problems.

Controlling comments can be difficult when there are a lot of community members, but the most common ways to get rid of them include either banning/blocking individual user accounts (and sometimes IP addresses altogether) or closing off comment sections entirely from a blog post, video page or topic thread [Internet troll] Media attention in recent years has equated trolling with online harassment. Troll detection assumes the availability of a signed social graph among users where signs indicate positive and negative relationships among users. The values u and v having relationship with one another. User u can have a positive and negative connection with user v .^[2] User generated content is fundamental to the notion of the "social web." However, an everyday issue facing technology companies and application developers is the encouragement of active user participation. One Barrier to participation is the presence of negative content -Messages that are not merely contentious, but angry, hostile, or abusive.^[3] In phase most users incline to be secular, others may engage in antisocial behavior, negatively affecting other users and harming the community. Such undesired behavior, which includes trolling, flaming, bullying, and harassment, is exacerbated by the fact that people swing to be less inhibited in their online interactions.^[4]

Bullying someone can also take place through the use of websites belonging to certain groups to effectively request the targeting of another individual or group. Examine forum user discussions of trolling, defined a troll as one who assemble the identity of genuinely wishing to be part of the group, while really aiming to cause disturbance for their own laughter. The definition can be broadened to include people who sincerely wish to be part of the group, but seek to influence the forum negatively, by continually starting arguments, criticizing or complaining. Internet communities who saturate personal trust, emotional commitments, and personal information, may find trolling particularly wounding, distressing and unsolvable.^[7]



MATERIALS AND METHODS

In Troll Vulnerability in Online Social Networks proposed model for troll vulnerability called as TVP (Troll Vulnerability Prediction) Problem. They assume that trolling occurs within an online user commitment ecosystem or environment, such as a social network, a microblogging system, or a discussion forum. Users subscribe content in the form of posts, and they interact with each other, generate discussions. We model interactions between posts as a directed graph $G = (V;E)$, where nodes $u \in V$ correspond to posts and there is an edge $(u; v)$, from post u to post v , if v is a reply to u . For Example, in Twitter, nodes may correspond to tweets and there is an edge from a tweet (node) u to all tweets (if any) that this tweet refers to. Similarly, in Facebook, nodes may correspond to comments on user posts.^[2]

In detecting trolls, they developed a general algorithm called TIA (Troll Identification Algorithm) to classify users of an online “signed” social network as poisonous (e.g. troll). In this paper the goal is to adjacent a single framework within which to identify venomous users. A major challenge in effectively identifying trolls is the fact that malicious users take a number of carefully designed steps that enable them to avoid detection. They proposed different graphs for decluttering operations that help simplify a large, convoluted SSN G into a smaller and simpler SSN G_0 . Intuitively, the idea is to remove some unwanted words they are searching in order to present their TIA algorithm with a simpler signed graph, empty of unconnected edges, that enables TIA to control more effectively. They tested all subsets of different decluttering operations and found the merger that yields the best results.^[4]

In Antisocial Behavior they characterize antisocial behavior in three large online discussion communities by analyzing users who were prohibited from these communities for online discussions. They use retroactive longitudinal analyses to quantify such behavior throughout an individual user’s holding in a community. This enables them to address various questions about antisocial Behavior: First, are there users that only become antisocial later in their community life? Second, does a community’s reaction to users’ antisocial behavior help them improve? Last, can antisocial users be effectively discover early on? To answer these questions, they examine three large online discussion-based communities: CNN.com, a general news site, Breitbart.com, a political news site, and IGN.com, a computer gaming site. They studied complete data from these websites: over 18 months, 1.7 million users contributed nearly 40 million posts and more than 100 million votes. In these communities, members that repeatedly violate community norms are eventually banned permanently. Such individuals are clear instances of antisocial users, and represent “ground truth” in their analyses.^[5]

In Identification Of Negative Comments On Internet Forum they use a machine learning technique - the Support Vector Machine (SVM) for sentiment classification. The objective of their paper is to build a real-time opinion mining framework which uses Web crawlers to keep searching for new reviews or comments. When the crawlers fetch new reviews or comments, the sentiment classifier they built before can analyze whether the reviews or comments are negative or not. The machine learning technique they used is the SVM. The SVM is to find a hyperplane in a high dimensional space and the hyperplane is used for distinguishing different classes. If a negative comment is found by the classifier, this comment will be reported to the enterprise. The enterprise can carefully examine the content of the negative comment and take proper actions immediately.^[6]

The negative comment identification system consists of three parts. The first part is a Web crawler which acquires the reviews and comments from Internet forums. The second part is the pre-processing module which transforms the reviews and comments into vectors suitable for the input of the classifier. The third part is the sentiment classifier using the SVM. Fig. 1 shows the flow chart of the whole process of the negative comment identification system.

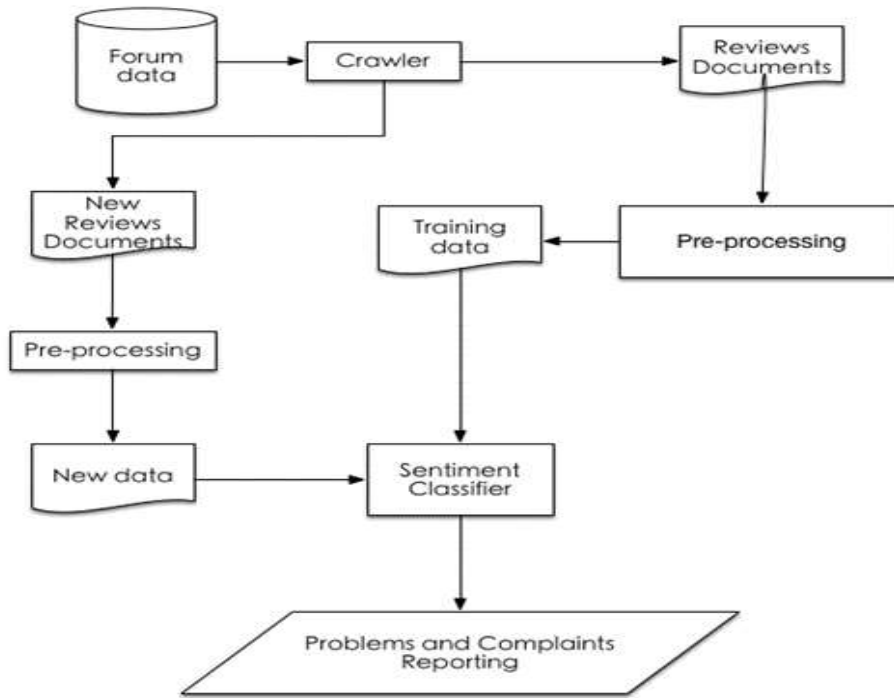


Fig. 1. Identification of Negative Comments

In the flow chart, the system first acquire the comments from Internet forums. Crawler programs are needed to collect the newly-added online reviews and comments automatically. Next, the reviews and comments are pre-processed. This phase includes sentence segmentation, word segmentation, stop words removal and ground truth labeling.^[7]

RESULTS ANALYSIS

Methods	Results	
TVP	Mean -- 0.76	Average – 0.92
TIA	MAP -- 0.5	Run Time – 0.4
CNN	Mean	Average
IGN	0.58	0.76
Breitbart		0.74
SVM	Negative Score – 0.86	Accuracy
	Non-Negative Score – 0.95	
		0.93
Troll –Trust Model	Accuracy For Different Datasets	
	Wikipedia – 0.89	
	Slashdot – 0.90	
	Epinions – 0.96	

CONCLUSION

Understanding and detecting trolling behavior in social networks has attracted considerable attention. In this paper, we took a different approaches focusing on trolling vulnerability, negative comments and algorithms to detect trolling behavior. We came across different approaches like TVP, SVM, TIA and Decluttering implemented in recent years. They have used different datasets from social community sites such as Reddit, Slashdot Zoo and CNN. Troll detection includes negative comments and positive comments separation and their impact. SVM gives overall accuracy of 0.93% which is better. So, by working on troll vulnerabilities and their causes we can take proactive measure to stop or minimize troll or bullying nature.

ACKNOWLEDGEMENTS

We are delighted to express our sentiments of gratitude to all who contribute their valuable guidance to us. We would like to express our admiration and thankful to Prof. R. R. Deshmukh, Head of Department, Department of Computer Science and Information Technology, Dr. BAMU, Aurangabad. We are also thankful to our guide Prof. S. N. Deshmukh, Department of Computer Science and Information Technology, Dr. BAMU, Aurangabad. We also thank the anonymous reviewers for their comments.

REFERENCES

- [1] Paraskevas Tsantarliotis, Evaggelia Pitoura and Panayiotis Tsaparas “Troll Vulnerability in Online Social Networks” in IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2016.
- [2] Srijan Kumar, Francesca Spezzano, and V.S. Subrahmanian, “Accurately Detecting Trolls in Slashdot Zoo via Decluttering” in IEEE, 2014.
- [3] Sara Owsley, Sood Elizabeth F. Churchill and Judd Antin , “Automatic Identification of personal insults on social news sites” in Journal of the Association for Information Science and Technology, October 2011.
- [4] Justin Cheng, Cristian Danescu-Niculescu-Mizily and Jure Leskovec “Antisocial Behavior in Online Discussion Communities” in arxiv.org/abs/1504.00680, April 2015.
- [5] Jung-Kuei Chang, Wei-Yun Hsu, Te-Chan Chen and Hui-Huang Hsu, “Identification of Negative Comments on Internet Forums” in 10th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, 2016.
- [6] Zhaoming Wu, Charu C. Aggarwal and Jimeng Sun, “The Troll Trust Model for Ranking in Signed Networks” in WSDM '16 Proceedings of the Ninth ACM International Conference on Web Search and Data Mining , February 22 - 25, 2016 .
- [7] Amy Binns, “DON'T FEED THE TROLLS!” in Journalism Practice Vol. 6 , Iss. 4,2012.

CITE AN ARTICLE

Tayade, P. M., & Shadulla, S. S. (2017). DISCOVER TROLLING IDENTITY IN SOCIAL MEDIA : REVIEW. INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY, 6(5), 359-363. doi:10.5281/zenodo.573510