Data Mining Social Media Model for Detecting Cybercriminal Network (DMSM-CN)

Ejiofor C. I
Department of Computer Science,
University of Port-Harcourt, Port-Harcourt, Nigeria

Onuodu .F
Department of Computer Science,
University of Port-Harcourt, Port-Harcourt, Nigeria

ABSTRACT

A Data Mining Social Media Model for Detecting Cybercriminal Network (DMSM-CN) has been designed. This model locate hidden pattern found in cybercriminal network with the aim of mitigating it. The model comprises of The model comprises of: Online social media, File and network analyzer, Positive data, Negative data, Classifier, Gibb sampling Algorithms, Laplacin Semantic Inference and inferential language modeling. The model will mitigate cybercrime, eliminating cyber network and enhancing the usefulness of social media.

Keywords

Cybercrime, cyber network, Social media

1. INTRODUCTION

Social media is largely implemented using computers, computer-peripherals and computer based networks for online socialization [8]. These components form a unified technological framework which facilitates the ease of sharing information, data and ideas [4]. Usually, user access social media services through web based platforms running on stationery and mobile devices which have been enhanced to accommodate smart devices [2]. These devices provide a universal pools for which users, individuals and communities can share, organize and coordinate relevant social media issues. Social media devices have indeed change the sphere of communication across the globe [5]. This communication have been provided at little or no cost to both parties; eliminating communication barriers and boundaries, providing remote and onsite access to relevant information and data. Social media has also provided a wide range of revenue both for individuals operating online businesses and for large organization struggling to open up its marketing base [5, 6]. Although social media has provided tremendous benefit, it has been misused by devious individuals or organization, heading for their selfish agendas. This has resulted in loss of personal information: credit card number, individual access identity and even personal authentication credentials. Social media misused has also created identity and privacy issues [3]. The huge amount of money lost annually due to the misuse of social media components is indeed tremendously running into billions of currencies globally [11].

Interestingly, data mining find it trace in social media through social media mining which is seen as the process or techniques of locating hidden patterns through the process of analyzing, representing, extracting patterns from data collected from wide range of social media activities with the aim of detecting similarity in patterns. Social media mining is largely implemented using tools for measuring, modeling and mining meaningful patterns from large-scale media [1].

Cybercrime implemented through cyber networks has been an area of interest to researchers and is growing tremendously [1].Cybercrimes has been associated repeatedly with these negative fraudulent issues emanating from cybercriminal operating cybercriminal networks. These networks have been used repeated in creating havoc for innocent populaces. This network are designed to mimic genuinely existing network with facilitates either expertly crafted or slightly differentiated causing serious confusion for unwiring victims [1, 9].

Therefore it is the intent of this research paper to design a Data Mining Social Media Model for Detecting Cybercriminal Network.

2. RELATED WORK

This section provide a brief review into several research on cybercrime, social media and cyber network with the hope of creating a research path for future study. The review was handled using a tabularization specifying the year of research, author, goal, strength, weakness and further studies. Tables 1 provide a brief review of related literatures pertaining to cybercrimes, cyber network and data mining.

Table 1: Review of Related Researches

SN	Author (Year)/Title	Goal	Strength/ Finding	Limitation/ Weakness	Further Research
1.	Neela, (2012), Impact of Cyber Crimes on Social Networking Pattern of Girls	The impact of Social media and cybercrime on girl	Cybercrime have made most girl alert	i. Purely investigative research ii. no model base solution	No Model base solution in creating
2.	Wajeb and Maha (2012) Cyber Threats In Social Networking Websites	Study the cyber threats in social networking Websites.	Classify and visualize cybercrime future trends	Investigative research which was not enhanced through model creation	No Model base solution in creating
3.	Reginald (2014) Social Media as a Channel and its Implications on Cyber Bullying	Elucidation of Social media and channel implication	Social media bring user into cyberspace.	Pure narrative research	No mathematical or model base solution.
4	Roshan and Afshar (2016) Issues and Challenges of Cyber Security for Social Networking Sites (Facebook)	Assessing Social Media using Facebook as case study	A comprehensive review of social media	Investigative research on primary on Facebook as a social media platform	No model base solution.

The research works captured on Table 1 shows clearly that most researches on cybercrime has focused mainly on investigative, narrative research which no precise solution designed or developed from a model based approach. This is necessary due to social media miming, highlighting the hidden pattern residual in social media data exemplified through cyber network. Therefore base on the limitation of these literatures; it is the intent of this research paper to provide a Data Mining Social Media Model for Detecting Cybercriminal Network (DMSM-CN)

3. DATA MINING SOCIAL MEDIA MODEL FOR DETECTING CYBERCRIMINAL NETWORK (DMSM-CN)

The designed model: Data Mining Social Media Model for Detecting Cybercriminal Network (DMSM-CN) is a text and semantics based mining model for detecting cybercriminal network. The model operates using several components integrated within the model. These components includes: online social media, file and network analyzer, positive data, negative data, training set, classifier, Gibbs sampling Algorithms, Laplacian semantic inference and inferential language modeling. Figure 1, graphically depicts the, Data Mining Social Media Model for Detecting Cybercriminal Network (DMSM-CN).

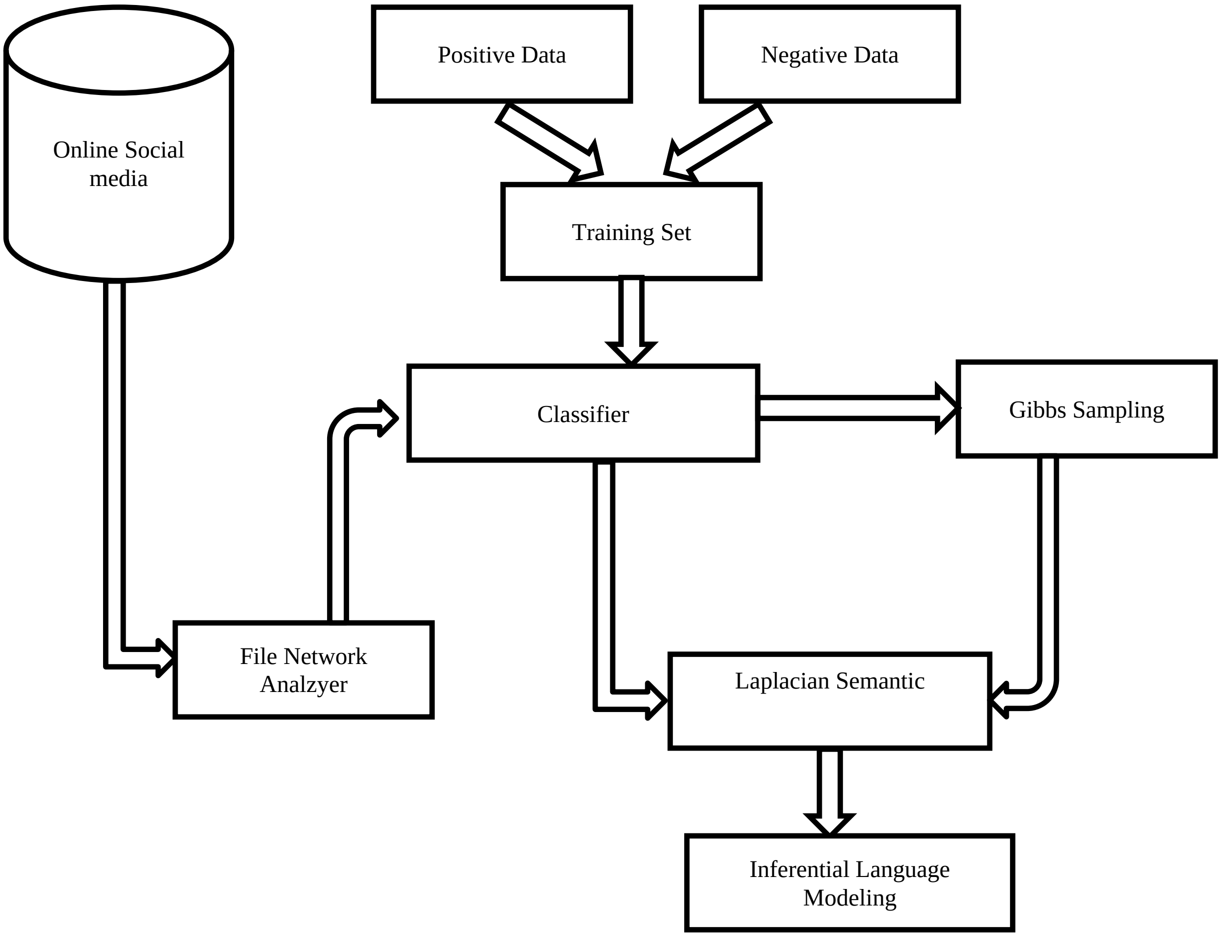

Figure 1: Data Mining Social Media Model for Detecting Cybercriminal Network (DMSM-CN).

The Data Mining Social Media Model for Detecting Cybercriminal Network (DMSM-CN) comprises of certain fundamental components which are elucidated as following:

i. **Online Social Media:** This module contains social media site where messages are extracted.

ii. **File and Network Analyzer:** This module contains the file analyzer and the network analyzer which extract messages from social media networks. The file analyzer extract messages from the social media sites while the network analyzer extracts the number of users connect by such message. The file analyzer keeps track of the extracted messages and the network analyzer keeps track of the users connected to such messages.

iii. **Positive Data:** The positive data contains words that could flag a cyber-criminal

iv. **Negative Data:** The negative data contains words that are not spiteful

v. **Training Set:** The module is where the naïve bayes is trained using the positive and negative data

vi. **Classifier**: The extracted files are passed to this module and are flagged either as malicious or not by the training set. This module classifies the extracted message. The file is classified using the naïve algorithm, and if a file is not flagged as containing malicious content it is discarded. The classifier classifies flagged documents either as labeled or unlabeled. Unlabelled messages are messages that refer to at least two users on the network.

vii. **Gibbs Sampling Algorithm**: The extracted unlabelled message is given to context sensitive Gibbs Sampling method for Latent topic modeling .The module uses the Gibbs sampling algorithm to infer a semantic label to the unlabeled messages which is known as latent concept. Semantic label for the unlabelled messages includes collaborative or transactional known as latent concept.

viii. **Laplacian Semantic Inference:** The latent concept and the labeled messages are sent to the Laplacian Semantic Inference. The Laplacian Semantic Inference determines the semantic content of the messages and rates the messages based on the semantic content. The latent concept and the labeled messages are sent to the Laplacian Semantic Inference. The Laplacian Semantic Inference determines the semantic content of the messages and this is used to rate the messages.

ix. **Inferential Language Modeling:** The rated messages are passed to the inferential language modeling module. This module generates the relationships between cyber criminals on the network.

4. MODEL DISCUSSION

Mitigating Cybercriminal network found across social media using data mining techniques is the focal of this research paper. The research paper has highlighted the fundamental security issues posed due to the propagation of social media. Notable among this issue is cybercriminal network. This research paper has provided a comprehensive framework in locating hidden patterns associated with cybercriminal network domiciled within user device and the explored network. The Model would be implemented with nay object oriented programming language: Java, C++, C#, and C. It is hoped; on full model implemented the following benefits will be derived:

a. Mitigating cybercrime

b. Eliminating cyber network

c. Enhancing the usefulness of social media

d. Providing a seamless integration between social media technologies.

5. CONCLUSIONS

This research paper has designed a concise model in addressing cybercriminal within social media. The designed model: Data Mining Social Media Model for Detecting Cybercriminal Network (DMSM-CN) was based on the identified limitation of previously researches tiled exclusively to narrative solution with no defined model. This research paper has provided a novel model with eight notable components: Online social media, File and network analyzer, Positive data, Negative data, Classifier, Gibb sampling Algorithms, Laplacin Semantic Inference and inferential language modeling. An architecture conceptualization was the intent of this research article with subsequent article geared toward it implementation.

6. REFERENCES

[1] Gladwell, M. (2011). "Malcolm Gladwell and Clay Shirky on Social Media and Revolution, Foreign Affairs March/April 2011". Foreignaffairs.com. Retrieved 24 April 2012.

[2] Hajirnis, A. (2015). "Social media networking: Parent guidance required". The Brown University Child and Adolescent Behavior Letter. 31 (12): 1–7.

[3] Hayat, T. Samuel-A., T. (2017). ""You too, Second Screeners?" Second Screeners' Echo Chambers During the 2016 U.S. Elections Primaries". Journal of Broadcasting & Electronic Media. 61 (2): 291–308.

[4] Kaplan A. M., Haenlein M. (2010). "Users of the world, unite, The challenges and opportunities of social media" (PDF). Business Horizons. 53 (1): 61. doi:10.1016/j.bushor.2009.09.003.

[5] Mangold, W. Glynn; F., David J. (2009). "Social media: The new hybrid element of the promotion mix". Business Horizons. 52 (4): 357–65.

[6] Mortimer, N. (2016) 'Magnum invests £13m in campaign to promote new Magnum Double' Available at: http://www.thedrum.com/news/2016/04/11/magnum-invests-13m-campaign-promote-new-magnum-double

[7] Neela M. M. (2012), Impact of Cyber Crimes on Social Networking Pattern of Girls, International Journal of Internet of Things 2012, 1(1): 9-15 DOI: 10.5923/j.ijit.20120101.02

[8] Obar, J. A.; Wildman, S. (2015). "Social media definition and the governance challenge: An introduction to the special issue". Telecommunications policy. 39 (9): 745–750.

[9] Pajala, M. (2012). "Television as an Archive of Memory?". Critical Studies in Television. **5** (2): 133–145.

[10] Reginald H. G. (2014), Social Media as a Channel and its Implications on Cyber Bullying, Presented at the DLSU Research Congress 2014

[11] Tang, Q.; Gu, B.; Whinston, A. B. (2012). "Content Contribution for Revenue Sharing and Reputation in Social Media: A Dynamic Structural Model". Journal of Management Information Systems. 29 (2): 41–75.