Sounds of Silence Breakers: Exploring Sexual Violence on Twitter

Aparup Khatua
WKW School of Comm. and Information
Nanyang Technological University
Singapore
akhatua@ntu.edu.sg

Erik Cambria
School of Comp. Sc. and Engineering
Nanyang Technological University
Singapore
cambria@ntu.edu.sg

Apalak Khatua
Strategic Management Area
XLRI - Xavier School of Management
Jamshedpur, India
apalak@xlri.ac.in

***Abstract*—Gender-based-violence is a serious concern in recent times. Due to the social stigma attached to these assaults, victims rarely come forward. Implementing policy measures to prevent sexual violence get constrained due to lack of crime statistics. However, the recent outcry on the Twitter platform allows us to address this concern. Sexual assaults occur at workplaces, public places, educational institutes and also at home. Policy level approaches and awareness campaign for these assaults would not be similar. So, we want to identify the risk factor associated with these sexual assaults. We extracted 0.7 million tweets during the *#MeToo* social media movement. Next, we employ deep learning techniques to classify these sexual violences. We observe that sexual assaults by a family member at own home is a more serious concern than harassment by a stranger at public places. This study reveals assaults by a known person are more prevalent than assaults by unknown strangers.**

***Index Terms*—Twitter, Sexual Violence, MeToo, Deep learning**

I. Introduction

Figures won't lie! 17,700,000 women are the victims of rape since 1998 in the United States. 321,500 victims go through the trauma of rape and sexual assault each year in the United States. 1 in 6 American women face rape or an attempt to rape in their lifetimes, and mostly they face these assaults during 12-34 years of their age [1]. Gender violence can have long-term detrimental effects on victims. Shockingly, 94% of women who have gone through these trauma experience symptoms of post-traumatic stress disorder [2]. More importantly, 33% (13%) of women who are raped contemplate (attempt) suicide [3]. These statistics are not only alarming but also disturbing. Sexual violence is becoming a serious concern in the society. The challenging aspect for policymakers is the social stigma attached to these assaults. Sexual violence 'has been traditionally brushed aside due to the fear of shame, retribution and retaliation' [4]. Thus, victims rarely come forward, and it leads to 'serious underreporting of rape crime statistics' [5]. Consequently, implementing appropriate policy measures get constrained due to lack of structured information.

The situation has drastically changed in October 2017. The tipping point was the public revelations of sexual harassment allegations against the Hollywood producer Harvey Weinstein. Actress Alyssa Milano triggered the movement by encouraging the victims to post *#MeToo* as a status update to show the magnitude of sexual assault. Influential Hollywood personalities, such as Angelina Jolie or Gwyneth, joined the movement. On Twitter platform victims had an 'opportunity without having to go into detail about their stories if they did not want to.' Some victims post *#MeToo* without details, whereas 'others opened up in wrenching detail about abuse they had never before shared in public' [6]. This leads to the rise of non-anonymous self-reported stories, which can be viewed as a social movement by women expressing anger about sexual violence [7]. These Silence Breakers were chosen as TIME's Person of the Year 2017.

IEEE/ACM ASONAM 2018, August 28-31, 2018, Barcelona, Spain
978-1-5386-6051-5/18/$31.00 © 2018 IEEE

This voluminous tweet data allowed us to probe risk factors associated with sexual violence. Sexual violence not only occurs at workplaces or public places but also at educational institutes as well as home. We extracted, using the Twitter API, 0.7 million tweets which had the hashtag *#MeToo* and employed deep learning based text classification techniques to investigate the extent of sexual violence across various locations. To the best of our knowledge, none of the prior studies has attempted to address this. On the contrary to our expectation, we observe that sexual assaults by a known person are more prevalent than assaults by unknown strangers. Our preliminary data analysis a priori suggests that 1 out of 4 sexual assaults are by close family members or friends which indicates the severity of the problem in our society, and the need for stringent policy measures.

II. Twitter, Sexual Violence & Prevention

Traditionally social activists used to protest in public places such as roads or public parks. However, in recent times activists are preferring digital platforms, such as Twitter, for various socio-political movements which range from *#BlackLivesMatter* [8] to *#Brexit* [9], [10]. Feminist movements are no different because Twitter provides 'an outlet for people to organize, voice, and connect their opinions, despite these individuals never meeting in person' [11]. Twitter not only allows its users to express their opinions by posting a tweet but also users can interact with others by retweeting or replying. Like-minded users can associate with each other by using a specific hashtag.

Recent *#MeToo* campaign, against sexual assaults and gender violence, demonstrates the power of social media. For instance, Microsoft has changed its sexual harassment policies as a response to *#MeToo* movement. To probe public opinion regarding gender-based-violence, [12] collected nearly 14 million tweets and argued that social media data can complement survey-based data collection for policy designing. Another study [5] probed whether social media behavior can predict rape and sexual abuse, and found uses of misogynistic language is an indicator of per capita rate of rape in the United States. Shockingly, another study noted that ‘users who engaged in victim blaming were more likely to be retweeted and have more followers than Twitter users who engaged in tweeting victim support content’ [13]. Thankfully, another study [4] found that tweets are ‘showing empathy and encouraging others to continue the *#MeToo* movement.’

In brief, sexual violence prevention requires a multipronged approach which ranges from influencing policies and legislation, changing organizational practices as well as strengthening individual knowledge and skills [14]. The Centers for Disease Control and Prevention (CDC) began addressing sexual violence in 2001 and CDC is trying to preventing sexual violence before it occurs. However, the ‘foundations of efforts for preventing sexual violence come from both the feminist movement and from empirical data from research on associated risk and protective factors’ [14].

This paper aims to address the second aspect. We are trying to probe different types of sexual violence and associated risks with them. The *#MeToo* campaign allows us to gather data about the offenders and locations of sexual violence. Identifying the vulnerable locations, from the perspective of preventing sexual violence, is important for framing appropriate policies and legislation. Workplace or public transport are widely recognized in the media, but it is also important to understand the domestic violence and sexual assaults by family members. Alarmingly our preliminary analysis reveals a significant portion of our tweets mentions about sexual violence by close family members such as stepfather, uncle, cousin brothers, mom’s boyfriend, and so on.

III. Methodology

We define our problem as a classification task. So, we aim to develop a framework to classify a tweet as: assault at the workplace by colleagues, assault at the educational institute by teachers or classmates, assault at public places by strangers, assault at home by a family member, multiple instances of assaults or a generic tweet about sexual violence. To prevent sexual violence before it occurs, it is important to gauge the severity and nature of risk factors associated with above typologies. However, manual way of labeling various types of assaults is not scalable. In the context of hate speech detection, prior studies noted that developing an automated framework for ‘classifying a tweet ... is quite challenging due to the inherent complexity of the natural language constructs’ [15]. Further, they pointed out that previous methods such as ‘manual feature extraction’ or using ‘representation learning methods followed by a linear classifier’ are inefficient in comparison to recent deep learning methods. Another study [16] also pointed out that ‘bag-of-words approaches tend to have high recall but lead to high rates of false positives’ because ‘lexical detection methods tend to have low precision because they classify all messages containing particular terms.’ Thus, prior studies which considered supervised learning has failed to distinguish between various typologies. Following this stream of research, we consider deep learning techniques to train a multi-class classifier for identifying different types of sexual assaults.

Table1: Sample Tweets from Gold Standard

Types (#)	**Tweets**
By Family and friends @home (292 tweets)	I was sexually abused by a family member as a child ... not believed by my dad which has caused much pain throughout my life #metoo
@Educational Institutes (273 tweets)	Physics teacher during my first year of university I was scared to speak out #metoo
@Public Places (201 tweets)	Ive been groped before by a stranger while trying to get to the bathroom in a crowded bar #metoo
@Workplace (194 tweets)	#metoo I had to quit an amazing job because of my pervy boss
Multiple Instances (124 tweets)	#metoo as a little kid I was molested by two family members ... later as I grew up I was sexually harrassed by men in social settings
Generic Tweets (120 tweets)	For everyone with a friend or family member who had the courage to make a #metoo post ... Let them know that you love them and support them

A. Data

As we mentioned earlier, actress Alyssa Milano triggered the social movement against sexual harassment on October 15, 2017, and she used the hashtag *#MeToo* to create awareness about the severity of the sexual violence. This particular hashtag was tweeted by other Tweeter users nearly a million times in next 48 hours. During the period October 15, 2017, to October 26, 2017, we have extracted 0.7 million tweets with the hashtag *#MeToo*. A few other prominent hashtags in our corpus were *#AlyssaMilano*, *#harassed*, *#metoocampaign* and so on.

B. Preparation of Gold Standard

We have referred social science research [17], specifically gender-based-violence literature, to identify different types of sexual violence and offenders. Following the extant literature, we manually annotated 4 different categories: sexual assaults at home by family members or friend, sexual assaults at educational institutes such as schools and colleges by teachers or classmates, sexual assaults at public places such as assaults by strangers in public transportation, and sexual assaults by colleagues and bosses at workplace. Moreover, while going through the wrenching detail about sexual abuse, we noted some tweets mention about repeated occurrences of sexual assaults by the same offender or different offenders. Finally, another set of users tweeted some generic comments to support the movement. Thus, our tweets are categorized into these six classes. A sample tweet from each category is reported in Table I. We have 1204 annotated tweets to train the classifier. This gold standard is comprised of tweets where annotators have unanimously agreed about the classification. Table 1 also reports the type wise distribution of different categories in our gold standard.

C. Deep Learning Techniques

This paper considers the following deep learning techniques: multilayer perceptron (MLP), convolutional neural network (CNN), long short-term memory (LSTM) and bidirectional LSTM (Bi-LSTM). MLP consists of multiple layers of simple neurons that interact using weighted connections. After the initial input layer, there are multiple hidden layers and finally an output layer at the top [18]. CNN is popular for performing NLP related task such as sentiment analysis, text classification, machine translation, summarization and so on [19]. In CNN, the first layer embeds words into low-dimensional vectors. The next convolution layer performs convolutions over the embedded word vectors using multiple filter sizes refer Fig 1. The next layer consists of max-pooling to provide a fixed dimension output. Dropout layer switch is the popular approach to stochastically off a fraction of neurons to prevent the over fitting. On contrary to other methods, RNN can use the previous information in the processing of the present computation. In other words, it tries to predict the last word. In additional to RNN architecture, LSTM consists three gates input, forget, and output gates [20], [21], [22]. LSTM calculates the hidden state by considering the combination of three gates written below.

$$x = \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix}$$
$$f_t = \sigma(W_f.x + b_f)$$
$$i_t = \sigma(W_i.x + b_i)$$
$$o_t = \sigma(W_o.x + b_o)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c.x + b_c)$$
$$h_t = o_t \odot \tanh(c_t)$$

Finally, Bi-LSTMs are an extension of traditional LSTMs that can improve model performance on sequence classification problems. In LSTM input sequence feed is only in the forward direction, however, in the bidirectional feed, it is a two-way process refer Fig 2. Thus, bidirectional setup helps the network to learn better in the context of the classification problem. More importantly, Bi-LSTM preserves the information not only from past sequence but also from the future sequence which is appropriate for the complex sequence.

IV. Findings

CNN is our best performing model, and it has an overall accuracy of 0.83, followed by LSTM (0.82), Bi-LSTM (0.81) and MLP (0.77). We use rmsprop for CNN, LSTM, and Bi-LSTM as our optimizer. We have performed training in batch sizes of 64 for all four models. We used 80% data for training (which includes 20% data for validation), and the remaining unexposed 20% data were considered for testing the model accuracy. Our settings for dropout layer are 0.5 and 0.2 for BI-LSTM and CNN respectively. We have considered softmax activation in our final classification layer to predict the final class. Results were similar across different batch sizes, epochs, dropout rates, embedding dimensions, and different kernel size for CNN. Thus, for the sake of brevity, we report the best results. Fig 3. reports the confusion matrix for all 4 models. In all 4 confusion matrix, most of the misclassification occurs for the class multiple instances and generic tweets. We carefully looked into the annotated tweets and it reveals a disturbing pattern. Generally, we observe multiple instances are repeated sexual assaults of victims by the same offender or family members. For instance, a tweets as follows: *‘I was sexually assaulted multiple times as a kid by moms boyfriends, one of my stepfathers, one boyfriend ... I still have scars #metoo’* indicates multiple instances of sexual assaults but mostly by family members and friends. Thus, our classifier is wrongly classifying it under the *Family & Friend* category.

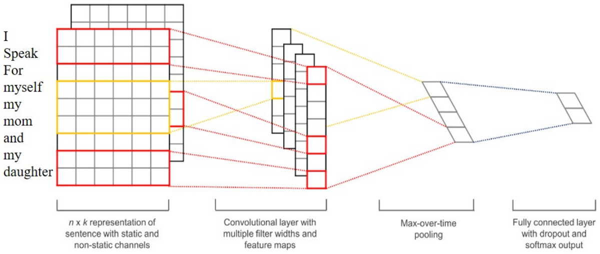

I
Speak
For
myself
my
mom
and
my
daughter
n x k representation of sentence with static and non-static channels
Convolutional layer with multiple filter widths and feature maps
Max-over-time pooling
Fully connected layer with dropout and softmax output

Fig. 1: CNN Architecture [23]

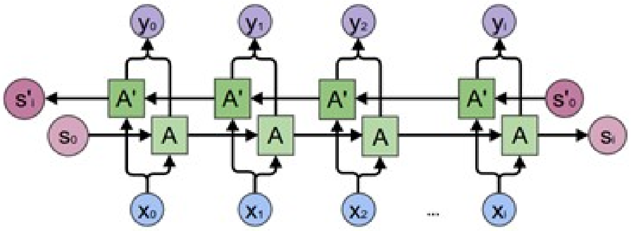
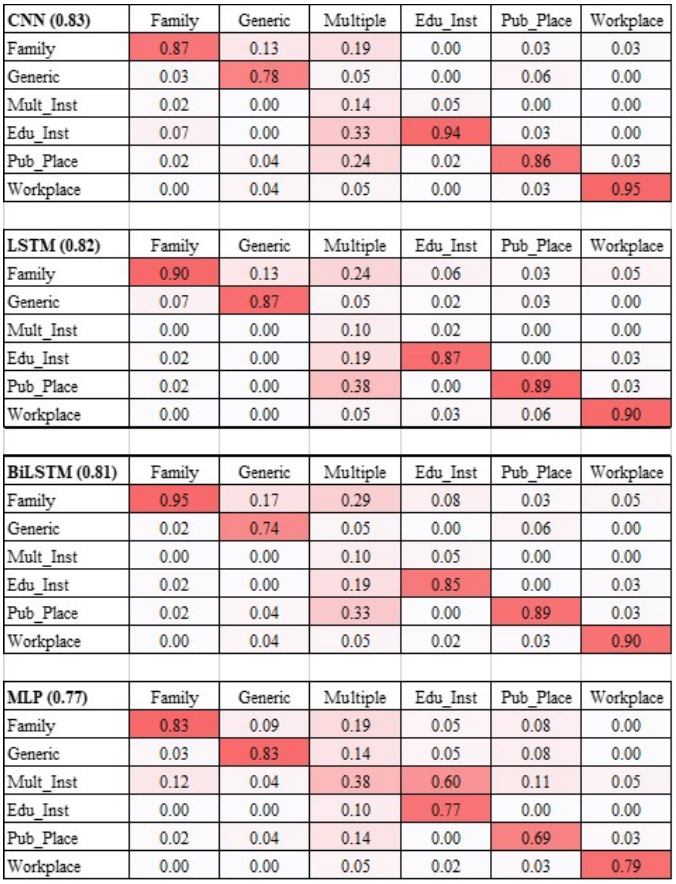


Fig. 2: General Structure of Bi-LSTM Source: colahs blog

V. Conclusion

In the wake of increasing gender-based-violence, it is important to probe the various risk factors associated with differentiate types of sexual assaults. *#MeToo* movements on Twitter platform allowed us to apply deep learning based lexical methods to classify sexual assaults in terms of locations and offenders. This approach can achieve relatively high accuracy at differentiating between different categories. However, careful analysis reveals that the presence of multiple instances within a tweet can hinder accurate classification. Future studies need to probe it further.



CNN (0.83)	Family	Generic	Multiple	Edu_Inst	Pub_Place	Workplace
Family	0.87	0.13	0.19	0.00	0.03	0.03
Generic	0.03	0.78	0.05	0.00	0.06	0.00
Mult_Inst	0.02	0.00	0.14	0.05	0.00	0.00
Edu_Inst	0.07	0.00	0.33	0.94	0.03	0.00
Pub_Place	0.02	0.04	0.24	0.02	0.86	0.03
Workplace	0.00	0.04	0.05	0.00	0.03	0.95

LSTM (0.82)	Family	Generic	Multiple	Edu_Inst	Pub_Place	Workplace
Family	0.90	0.13	0.24	0.06	0.03	0.05
Generic	0.07	0.87	0.05	0.02	0.03	0.00
Mult_Inst	0.00	0.00	0.10	0.02	0.00	0.00
Edu_Inst	0.02	0.00	0.19	0.87	0.00	0.03
Pub_Place	0.02	0.00	0.38	0.00	0.89	0.03
Workplace	0.00	0.00	0.05	0.03	0.06	0.90

BiLSTM (0.81)	Family	Generic	Multiple	Edu_Inst	Pub_Place	Workplace
Family	0.95	0.17	0.29	0.08	0.03	0.05
Generic	0.02	0.74	0.05	0.00	0.06	0.00
Mult_Inst	0.00	0.00	0.10	0.05	0.00	0.00
Edu_Inst	0.02	0.00	0.19	0.85	0.00	0.03
Pub_Place	0.02	0.04	0.33	0.00	0.89	0.03
Workplace	0.00	0.04	0.05	0.02	0.03	0.90

MLP (0.77)	Family	Generic	Multiple	Edu_Inst	Pub_Place	Workplace
Family	0.83	0.09	0.19	0.05	0.08	0.00
Generic	0.03	0.83	0.14	0.05	0.08	0.00
Mult_Inst	0.12	0.04	0.38	0.60	0.11	0.05
Edu_Inst	0.00	0.00	0.10	0.77	0.00	0.00
Pub_Place	0.02	0.04	0.14	0.00	0.69	0.03
Workplace	0.00	0.00	0.05	0.02	0.03	0.79

Fig. 3: CNN, LSTM, Bi-LSTM & MLP Confusion Matrix

References

[1] A. Vagianos, "30 alarming statistics that show the reality of sexual violence in america," *Huffington Post*, April-05-2017.

[2] B. O. Rothbaum, E. B. Foa, D. S. Riggs, T. Murdock, and W. Walsh, "A prospective examination of post-traumatic stress disorder in rape victims," *Journal of Traumatic stress*, vol. 5, no. 3, pp. 455–475, 1992.

[3] D. G. Kilpatrick, "Rape in america: A report to the nation," *Technical Report*, 1992.

[4] L. Manikonda, G. Beigi, H. Liu, and S. Kambhampati, "Twitter for sparking a movement, reddit for sharing the moment:# metoo through the lens of social media," *arXiv preprint arXiv:1803.08022*, 2018.

[5] R. Fulper, G. L. Ciampaglia, E. Ferrara, Y. Ahn, A. Flammini, F. Menczer, B. Lewis, and K. Rowe, "Misogynistic language on twitter and sexual violence," in *Proceedings of the ACM Web Science Workshop on Computational Approaches to Social Modeling (ChASM)*, 2014.

[6] CBS/AP, "More than 12m "me too" facebook posts, comments, reactions in 24 hours," *CBS News*, October-17-2017.

[7] M. ElSherief, E. M. Belding, and D. Nguyen, "# notokay: Understanding gender-based violence in social media." in *ICWSM*, 2017, pp. 52–61.

[8] M. De Choudhury, S. Jhaver, B. Sugar, and I. Weber, "Social media participation in an activist movement for racial equality." in *ICWSM*, 2016, pp. 92–101.

[9] A. Khatua and A. Khatua, "Leave or remain? deciphering brexit deliberations on twitter," in *Data Mining Workshops (ICDMW), 2016 IEEE 16th International Conference on*. IEEE, 2016, pp. 428–433.

[10] A. Khatua, E. Cambria, A. Khatua, and I. Chaturvedi, "Let's chat about brexit! a politically-sensitive dialog system based on twitter data," in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2017, pp. 393–398.

[11] M. K. Maas, H. L. McCauley, A. E. Bonomi, and S. G. Leija, ""i was grabbed by my pussy and its# notokay": A twitter backlash against donald trump's degrading commentary," *Violence Against Women*, p. 1077801217743340, 2018.

[12] H. Purohit, T. Banerjee, A. Hampton, V. L. Shalin, N. Bhandutia, and A. P. Sheth, "Gender-based violence in 140 characters or fewer: A# bigdata case study of twitter," *arXiv preprint arXiv:1503.02086*, 2015.

[13] M. Stubbs-Richardson, N. E. Rader, and A. G. Cosby, "Tweeting rape culture: Examining portrayals of victim blaming in discussions of sexual assault cases on twitter," *Feminism & Psychology*, vol. 28, no. 1, pp. 90–108, 2018.

[14] D. S. Lee, L. Guy, B. Perry, C. K. Sniffen, and S. A. Mixson, "Sexual violence prevention," *The Prevention Researcher*, vol. 14, no. 2, pp. 15–20, 2007.

[15] P. Badjatiya, S. Gupta, M. Gupta, and V. Varma, "Deep learning for hate speech detection in tweets," in *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 2017, pp. 759–760.

[16] T. Davidson, D. Warmsley, M. Macy, and I. Weber, "Automated hate speech detection and the problem of offensive language," *arXiv preprint arXiv:1703.04009*, 2017.

[17] D. Scully and J. Marolla, "Convicted rapists' vocabulary of motive: Excuses and justifications," *Social problems*, vol. 31, no. 5, pp. 530–544, 1984.

[18] S. K. Pal and S. Mitra, "Multilayer perceptron, fuzzy sets, classifiaction," 1992.

[19] W. Yin, K. Kann, M. Yu, and H. Schütze, "Comparative study of cnn and rnn for natural language processing," *arXiv preprint arXiv:1702.01923*, 2017.

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[21] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," 1999.

[22] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *IEEE Computational Intelligence Magazine*, vol. 13, no. 3, pp. 55–75, 2018.

[23] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.