Detecting Radical Text over Online Media using Deep Learning

Armaan Kaur  
armaankaurgill@gmail.com  
Cyber Security Research Centre,  
Punjab Engineering College, (Deemed To Be University)  
Chandigarh, India

Jaspal Kaur Saini  
sainijass87@gmail.com  
Cyber Security Research Centre, Punjab Engineering College, (Deemed To Be University)  
Chandigarh, India

Divya Bansal  
divya@pec.edu.in  
Cyber Security Research Centre, Punjab Engineering College, (Deemed To Be University)  
Chandigarh, India

ABSTRACT

Social Media has influenced the way people socially connect, interact and opinionize. The growth in technology has enhanced communication and dissemination of information. Unfortunately, many terror groups like jihadist communities have started consolidating a virtual community online for various purposes such as recruitment, online donations, targeting youth online and spread of extremist ideologies. Everyday a large number of articles, tweets, posts, posters, blogs, comments, views and news are posted online without a check which in turn imposes a threat to the security of any nation. However, different agencies are working on getting down this radical content from various online social media platforms. The aim of our paper is to utilise deep learning algorithm in detection of radicalization contrary to the existing works based on machine learning algorithms. An LSTM based deep forward neural network is employed to detect radical content. We collected total 61601 records from various online sources constituting news, articles and blogs. These records are annotated by domain experts into three categories: Radical(R), Non-Radical (NR) and Irrelevant (I) which are further applied to LSTM based network to classify radical content. A precision of 85.9% has been achieved with the proposed approach.

CCS CONCEPTS

• Information systems → Data analytics; Social networking sites;  
• Computing methodologies → Neural networks.

KEYWORDS

Social Media, Radicalisation, Deep Learning, Radical Content, LSTM

ACM Reference Format:

Armaan Kaur, Jaspal Kaur Saini, and Divya Bansal. Detecting Radical Text over Online Media using Deep Learning. In , ACM, New York, NY, USA, 8 pages.

1 INTRODUCTION

Different social media platforms follow different approach of connecting people and sharing information. With time, social media has penetrated and integrated into our lives so well that we share all the personal data and check-ins online. Such is the extent that it has opportune influence the opinions and decision-making of consumers. Reachability to masses has been expedited by the advancement in Internet facilities. It is startling how people leave word about their daily activities using social media and share information that actually affects the users connected. Hence, users with similar intents come aboard online.

A major drawback of such advancements is spreading of radicalization has eventually migrated from physical interactions at public places to virtual conversations over social media platforms. At present, the Islamic State (IS) is indisputably the leading organization in spreading of their propaganda online. The extremist groups spread their propaganda online, influence vulnerable frustrated individuals, motivate towards their ideology, recruit online, raise funds from supporters, instigate people to become members of aggressive communities and provoke lone-wolf terrorist into acts of violence. Based on the studies reviewed by Hassan et al. [8], contingent evidences have been provided in favour of fact that exposure of a person to radical content poses higher risks of engaging in political violence. Active seekers are at greater threat than passive ones.

Doosje et al. defines radicalization as: "Radicalization is a process through which people become increasingly motivated to use violent means against members of an out-group or symbolic targets to achieve behavioral change and political goals."[3] The model of radicalization is summarized in [3] undergoing three phases: (1) Sensitive phase: Being sensitive to radical ideas. (2) Membership phase: Becoming part of a radical group. (3) Action phase: Act on behalf of the ideologies of the radical group. Apart from recruiting attackers, intermediaries are also recruited online for carrying messages and spreading martyr videos. The major threats due to widespread of extremist ideas over the internet are enlisted below:

- Recruitment of ISIS fighters and searching for lone wolf terrorists online.
- Rapid dissemination of extremist ideologies.
- Identification of outraged, unemployed, anti-nationalist youth online that can be easily influenced.
- Many other jihadi groups have started adopting ISIS’s method of mobilizing supporters globally through social media.
- The establishment of online jihadi community may accelerate more rapidly with increasing global radicalization.
- Greater penetration of terror groups into social media accounts to target youth.
- Online donations and fund raising activities are becoming more digital and ease to reach masses over social media makes such activities spread easily.
The increasing role of Internet in contemporary terrorism has persuaded the world leaders and governments to call on social media companies to put more efforts in getting rid of violent and extremist content floating on their platforms. This would further hamper freedom of speech, therefore targeted deletion requires an effective analysis of content posted online. With the upsurge in development of radicalization, it is important to devise ways for detection and keep a technical edge with advancement in technologies. Also, Mukherjee et al. [14] refers to the limited research in the area of exploring texts of discussion forums deeply. As many such platform are used for contentious issues such as political, social and religious, hence these are topics of discussion and often go unaddressed by governments but it is critical to recognise them. Topic modelling using maximum entropy is one of the text analytics technique utilised in previous work [14] and compared with LDA and SLDA.

Our work proposes a novel approach in detection of radicalization on online media that put to use Deep Neural Networks using word embeddings. Exploring the field of Deep Learning is worthwhile that can certainly automate the process of extraction of complex data representations. These algorithms have a layered and hierarchical structure for learning and representation of data, where low level attributes are used to define high level attributes. [16]. Mudgal et al. apply entity matching with deep learning to look for the prospects of this much less explored area and finds out that Deep Learning has convincing results. It can easily outperform in the absence of data cleaning and therefore, push boundaries of current automated tasks [15]. The emerging field of deep learning calls for extensive research [21].

This paper presents an automated approach to detect radicalization over online social media. We utilised text analytics and deep learning mechanism to identify radical content. We started with studying various research evidences which throws light on application domains of deep learning techniques in text analytics. A comparative study of work done on multiple datasets is presented in current work. LSTM based feed forward neural network is employed to detect radical content. Further sections discuss the key details of our work.

2 RELATED WORK

For the purpose of literature survey, different papers were studied to have an insight into development of radicalization, techniques used to detect it and varied deep learning algorithms applied onto text analytics.

The comparison in Table 1 provides overview of techniques used for data radicalization along with the precision score achieved. As can be implied from Table 1, twitter can be used as a source for text analytics if annotated well and has been widely used in most of the studies. Fernandez et al. [4] analysed 114k tweets (17k from pro-ISIS users and 97k from general ones) and achieved an improvement of 4.5% with context modelling over keyword-based search. Mahmoud Barhamgi et al. [2] collected twitter messages exchanged between 20 accounts tagged radical by Kaggle dataset that had total 2317 messages both radical and neutral posted by selected users. Swati Agarwal et al. [1] combined datasets: UDI TwitterCrawl (Aug2012) and ATM TwitterCrawl (Aug2013), followed by filtration using language detection library to be left with 45.3 million english tweets.

| Study | Technique | Source of data | Precision |
|-------|-----------|----------------|-----------|
| Fernandez et al. (2018) | Context modelling based on ontologies and knowledge-bases, NB, SVM, J48 | Twitter | 0.859 |
| Barhamgi et al. (2018) |Ontology-based semantic description | Twitter | 0.77(Ontology), 0.38(Baseline) |
| Njagi Dennis Gitari et al. (2015) |lexicon-based approach | web forums, blogs, comment section of news reviews | Subjective sentence: 67.21(sem), 71.22(sem+hate), 73.42(sem+hate+theme) Without subjective sentence: 58.42(sem), 63.24(sem+hate), 65.32(sem+hate+theme) |
| Shwati Agarwal et al. (2015) | SVM, KNN | Twitter | 0.48(KNN), 0.78(SVM) |
| M. Munezero et al. (2013) | Bag-of-words, Ontology-based emotion description, SVM, Multinomial NB, J48(Decision tree) | Movie review, ISEAR, wikipedia article | SVM: 0.891(ASB+ISEAR), 0.955(ASB+MovRev), 0.883(ASB+Wiki), J48: 0.832(ASB+ISEAR), 0.966(ASB+MovRev), 0.821(ASB+Wiki), MNB: 0.896(ASB+ISEAR), 0.996(ASB+MovRev), 0.908(ASB+Wiki) |
| Pir Abdul Rasool Qureshi et al. (2011) | DETECT framework | blogs, discussion forums, news articles, etc | NA |

Table 1: Comparison of studies based on their techniques, sources of data and accuracies.

The tweets were manually annotated as hate and extremism promoting tweets and semi-supervised learning followed as the learning was performed by training the model over data of one class only. Overall LibSVM can be found to outperform KNN.

The classification model can also be trained over dataset collected from multiple sources like blogs, discussion forums, news and articles as all these sources have opinions of users. Apache Nutch has been extended into framework proposed in [17] for acquisition of open source information like blogs, discussion forums, news articles, facebook, data provided by analysts, etc. Njagi Dennis Gitari et al. [6] has chosen web forums, blogs and comment sections of news reviews as source of data. The corpus is divided into two: first one consisted of 30 blogs and second one consisted of...
150 pages having paragraphs from documents, labelled as strongly, weakly or not hateful. The classification is based on a lexicon built using semantic, hate and theme-based features from dictionary and corpus both. The precision improved with addition of hate and thematic elements to semantics and also further with the use of subjective sentences. The results for strongly hateful sentences are shown in the table.

M. Munezero et al. [15] collected corpora from antisocial behaviour texts (ASB), Movie Reviews and Wikipedia extracts. ISEAR corpus used for emotion description. WEKA tool is used for the implementation of machine learning algorithms: SVM, MNB, and J48 decision trees. In most of the cases, the precision can be seen to have improved with addition of emotions. The MNB has been found to have performed better compared with SVM and J48 when bag-of-words features combined with emotions. The addition of sentiment analysis to extend the work has been suggested.

Table 1 presents various techniques used for detecting radicalization. It is seen from the table that deep learning algorithms has not been utilised in the field so far. As can be seen, by and large the performance of classification algorithms depends upon goodness of data representation. The task of feature engineering consumes a lot of effort and directly impacts the performance of machine learning algorithm. Deep learning is one such area of research that has a lot of effort and directly impacts the performance of machine learning algorithm. Deep learning is one such area of research that has been explored for text analytics domain studied is shown in Table 2.

| Study            | Technique                          | Dataset                                      |
|------------------|------------------------------------|----------------------------------------------|
| Zhou et al. (2015) | C-LSTM Neural Network          | Stanford Sentiment Treebank (SST) consisting of movie reviews |
| Zhou et al. (2016) | BLSTM-2DPooling on 6 different classification tasks | Stanford Sentiment Treebank dataset            |
| Xu et al. (2017)  | Naive Bayes, LSTM and Social Network Analysis | Twitter data of a terrorist organization and supporting communities |
| Rybinski et al. (2018) | Logistics regression using bag-of-words and n-gram schemes compared with DNN operating on word embeddings | Toxic Comment Classification Challenge dataset |
| Tommasel et al. (2018) | SVM and RNN to analyze different feature sets | Dataset presented in Kumar et al. [10] |

Table 2: Comparison of studies on text analytics using deep learning on the basis of technique and dataset used.

The proposed architecture is as shown in Figure 1, divided into two parts: (1) Data preparation comprising of data collection, annotation and cleaning the raw data. (2) Training and Visualisation comprising of generating word embeddings, then classifying the texts as radical or non-radical, followed by trend analysis. The details of each component is described in detail in the following sections:

3 PROPOSED WORK

3.1 Data Preparation

(1) Data Collection

The data is collected from multiple sources namely: Blogs, Articles, Twitter, News. We have taken expert guidance to identify data sources that may contain radical content. Details of the same is shown in Table 3 below. We developed crawler that collected 61601 records from online social media sources.

(2) Data Annotation

For the purpose of data pre-processing and annotations, we defined radicalization as a general term for the online activities and discussions over online social media which support terrorism in the Indian subcontinent by the name of Jihad/Islam. It can be seen as course of action that provokes
vulnerable users towards violent extremism and propagating anti-nationalism which is against the interests of the nation. The various contours of radicalization can be summarized as:

- Abusing the Indian Army Troops, politicians and using insects as metaphors for them.
- Anti-national speeches with trigger factors
- Justification of terrorism by the name of Islam or Jihadism or Mujahideen

Table 3: Details of data collected.

| Type  | Name                      | Time Period        | Records |
|-------|---------------------------|--------------------|---------|
| News  | Greater Kashmir           | 2018               | 52      |
|       | Kashmir Reader            | 2014-2018          | 35637   |
|       | Kashmir Monitor           | 2018               | 1403    |
|       | Tribune                   | 2016-2018          | 20994   |
| Articles | Huriyat                  | 2010-2013          | 62      |
|       | Huriyat                   | 2017-2019          | 591     |
| Blogs | JKLF World                | 2009-2018          | 478     |
|       | JKPFL                     | 2010-2015          | 84      |
|       | Kashmir Truth be Told     | 2006-2010          | 145     |
|       | Peoples League            | 2009-2016          | 276     |
|       | Sameer Bhatt              | 2006-2017          | 475     |
|       | Tanveer and Kashmir       | 2008-2019          | 1343    |
|       | United Kashmir            | 2008-2014          | 61      |
| Total Records |                               |                    | 61601   |

Table 4: Sample annotations without preprocessing

| Sample Text | Label |
|-------------|-------|
| Srinagar: All Parties Hurriyat Conference, Chairman Syed Ali Geelani On Saturday Said That Forced And Military Occupation Of India Is The Root Cause Of All Miseries And Problems And Added That Authorities In New Delhi Always Applied Its Shrived Tactics To Supress Our Sentiments. | R |
| MUZAFFARABAD: United Jehad Council Has Said That It Has Full Faith And Confidence On Joint Resistance Leadership And Venerates Them. In A Statement To Media, Syed Sadakat Hussain, The Spokesperson Of United Jehad Council Said A Confusion And Misunderstanding Were Being Advertently Or Inadvertently Created After UJC Under A Well Chalked Out Strategy Issued A Schedule From July 8 To 13. "Now UJC Has Decided To Restrict This Schedule To Base Camp Only While On Ground In Kashmir, Joint Resistance Leadership Will Take A Final Call," The Spokesperson Said. | NR |
| No more comment moderation, Dear friends,. I will be away for a few months because of some issues that need my attention. I have therefore removed the requirement for comment moderation. All posts will appear instantaneously. The only difference is that "anonymous" posts are not possible now. Any registered user (Gmail, Google, OpenID, facebook, wordpress, blogger etc) will be able to post instantaneously. | I |

- Terrorists who died while fighting or got assassinated being termed as martyrs and said to have sacrificed their lives

Table 3: Details of data collected.
Detecting Radical Text over Online Media using Deep Learning

• Motivating readers to support extremism

Some sample annotations are enlisted in the Table 4. We provided collected records to two domain experts with definitions of radicalisation and requested them to annotate the given records with following instructions:
(a) You are provided data from news, articles, blogs and discussions.
(b) Read it carefully and label the data into three categories:
   1. Radical (R) 2. Non-Radical (NR) 3. Irrelevant (I)
(c) Label the data as Radical (R): Based on the presence of above-mentioned features of radicalisation devised for the purpose of annotation, label the data as Radical (R).
(d) Label the data as Non-Radical (NR): All posts which are relevant to the topic of radicalisation but fails to meet the requirements of our definition mentioned above i.e. the post which does not provoke or motivate or justify the topic of radicalisation will be labelled as (NR). For example: related facts about any such radical event like reporting of any event, assassination, arrest or violent killings etc that talks about nation, government, politics, terrorism and conflicted areas in Indian interests.
(e) Label the data as Irrelevant (I): Label the data as neutral that is completely off the topic or including advertisement.

3.2 Cohen’s Kappa Coefficient

Cohen’s Kappa is a measure of inter-expert agreement between two experts/raters on our annotated data. Cohen’s kappa assesses the agreement between two raters classifying items(n) into mutually exclusive classes(c). Cohen’s kappa coefficient (κ) determines inter-rater agreement for qualitative items that can be categorized. It is calculated as:

\[
\kappa = \frac{p_o - p_e}{1 - p_e}
\]

where,

\[p_o\] can be termed as relative observed agreement among experts (similar to accuracy),
\[p_e\] can be termed as hypothetical probability of chance agreement,

using the data observed for calculating the probabilities of each observer by randomly seeing each category.

| Subjects | 1274 |
|----------|------|
| Experts  | 2    |
| Kappa coefficient | 0.796 |

Table 5: Kappa Coefficient values.

The value for \( \kappa \) comes out to be 0.796 which is a sign of substantial agreement among two raters as shown in Table 5. Considering the significant value of kappa coefficient, we further moved on to utilising this annotated data to detect radicalisation using our proposed approach.

3.3 Training and Visualization

(1) Generating Word Embeddings

All the word embedding representation methods are based on a hypothesis that words generally carry similar meaning that occur in similar contexts. [7] A good representation of text should aim at maintaining the analogy between words during course of word embedding generation. Here, the pre-processed data is first converted to vectors with an ASCII supported integration.

The dense vector representation of words learned by word2vec have shown remarkable progression in carrying semantic meanings [12]. Apparently, word2vec framework has been employed and vectors are neutrally trained separately for Radical and Non-Radical texts to generate word embeddings employing a Levenberg training mechanism. The validation criteria used while training: Propogative iterations and Gradient satisfaction. The weights are updated according to the equation:

\[ w_o = w_i \ast \alpha + b \]

(2)

(3) LSTM and Fully-connected Layer

A feedforward neural network supported by LSTM is used for weight training as it is presumes texts to be interrelated. Many models use word embeddings as direct features while some use avg task i.e. calculate the weighted average of word embeddings in similar context as representations of text[11]. Hence, the fully connected layer predicts labels for test datasets by comparing thresholds, generated using mean of radical and non-radical weights.

(3) Trend Analysis

The results of classification are visualized graphically which are discussed in detail in next section.

4 EXPERIMENTAL RESULTS

We have choosen data from articles and blogs as they were identified radical by associated experts in the field of information warfare and therefore prodigiously contain radical content. There are total 1274
records in all after data cleaning where each record is comprised of 500 words on an average. The radical texts from the dataset are plotted over a timeline starting from 2006 till 2018 as shown in the Figure 2. It can be drawn that blog JKLF have been posting radical content for longer time than any other source, and Hizbul have the maximum radical data that have been posted in a year.

The results of the classification algorithm applied on different training set proportions can be seen in the Figure 3 which implies maximum accuracy of 73.44% is achieved with a 80:20 training and testing dataset ratio. The model starts to overfit beyond this ratio and accuracy drops as shown in Figure 3.

As can be seen, the value of kappa coefficient achieved (here 0.796) indicates the level of agreement between two annotators in terms of data of documents collected. However, some authors of the blogs have mentioned their opinions from both perspectives: in favor of and against the radical ideology. Such texts lead to the computation of mean which may occur closer to any of the radical or nonradical means with slight changes in text. In such cases, the prediction model may go inaccurate and hence the accuracy drops.

Further, the proposed approach is exercised over a 3-class classification for detection of irrelevant content in the data but accuracy declines to 34.23% due to the presence of varied heterogeneous data in the irrelevant tagged section of data. The mean square error is plotted for radical content prediction while testing the model. The Figure 4 exhibits the variations in prediction of radical texts on calculation of mean.
The proposed model is compared with few existing techniques for text analytics in order to establish the credibility of model in terms of different performance measures observed. At first, bag-of-words technique is employed to build a term document matrix. Then three machine learning algorithms namely SVM, Random Forest and MaxEnt are implemented to achieve classification of texts being radical or not radical. The algorithms applied have different ways of operating. Support Vector Machine focuses on finding a hyperplane in N-dimensional space that classifies the data points. Random forest takes into account many decision trees predicting classes individually, then the votes from all the trees are aggregated to decide final prediction of model. Maximum Entropy Classifier selects the model with largest entropy among all the models fitting training data for classification. The various performance measures being evaluated for comparison are enlisted in the Table 6.

| Approach              | Precision | Recall | F-score |
|-----------------------|-----------|--------|---------|
| Feedforward NN        | 85.96     | 53.26  | 65.77   |
| supported by LSTM     |           |        |         |
| Random Forest         | 73.50     | 64.00  | 65.55   |
| SVM                   | 53.50     | 30.50  | 45.00   |
| MaxEnt                | 69.50     | 68.00  | 68.55   |

Table 6: Different performance measures of our implementation expressed in percentage.

The comparison of precision score of different techniques is presented graphically in the Figure 5 to demonstrate a contrasting view of implemented algorithms.

![Figure 5: Comparison of precision score of different algorithms applied to the dataset.](image)

5 CONCLUSION AND FUTURE WORK

The radical data detection task has been carried out in numerous studies in the past ranging from ontology-based techniques [4] [15], to machine learning algorithms. The feature extraction task before applying classification methods has utmost importance and requires plenty of effort. Even then it might miss out on capturing semantics of the text. Deep learning can be applied to overcome the drawbacks. In the present work, we have utilised LSTM based approach to identify radical content over online media. This model can help online social media sites in omission of inappropriate content. The proposed approach has been able to overtstep most of the state-of-art techniques in terms of precision (85.9%). This model can be extended with additional layer of CNN for explicit identification of features.

6 ACKNOWLEDGEMENTS

This work was partially supported by Cyber Security Research Centre, Punjab Engineering College (Deemed to be University), Chandigarh, India. The author Jaspal K Saini is grateful to Visvesvaraya PhD scheme for Electronics and IT for funding this research.

REFERENCES

[1] Swati Agarwal and Ashish Sureka. 2015. Using knn and svm based one-class classifier for detecting online radicalization on twitter. In International Conference on Distributed Computing and Internet Technology. Springer, 431–442.
[2] Mahmoud Barhamgi, Abir Masmoudi, Raul Lara-Cabrera, and David Camacho. 2018. Social networks data analysis with semantics: application to the radicalization problem. Journal of Ambient Intelligence and Humanized Computing (2018), 1–15.
[3] Bertjan Doosje, Fathali M Moghaddam, Arie W Kruglanski, Arjan De Wolf, Laesbeth Mann, and Allard R Feddes. 2016. Terrorism, radicalization and de-radicalization. Current Opinion in Psychology 11 (2016), 79–84.
[4] Miriam Fernandez and Harith Alani. 2018. Contextual Semantics for Radicalisation Detection on Twitter. (2018).
[5] Miriam Fernandez, Moziah Asif, and Harith Alani. 2018. Understanding the roots of radicalisation on twitter. In Proceedings of the 10th ACM Conference on Web Science. ACM, 1–10.
[6] Njagi Dennis Gitari, Zhang Zuping, Hanyurwimfura Damien, and Jun Long. 2015. A lexicon-based approach for hate speech detection. International Journal of Multimedia and Ubiquitous Engineering 10, 4 (2015), 215–230.
[7] Zellig S Harris. 1954. Distributional structure. Word 10, 2-3 (1954), 146–162.
[8] Ghayda Hassan, Sébastien Brouillette-Alarie, Séraphein Alava, Divina Frau-Meigs, Lysiane Lavanie, Arber Fetiu, Wynnapaul Varela, Evgenii Borokhovski, Vivek Venkatesh, Rouseau Cécile, et al. 2018. Exposure to extremist online content could lead to violent radicalization: a systematic review of empirical evidence. International journal of developmental science Preprint (2018), 1–18.
[9] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. 2007. Why we twitter: understanding microblogging usage and communities. In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis. ACM, 56–65.
[10] Ritesh Kumar, Aishwarya N Reganti, Akshit Bhatia, and Tushar Maheshwari. 2018. Aggression-annotated corpus of hindi-english code-mixed data. arXiv preprint arXiv:1803.09402 (2018).
[11] Siwei Lai, Kang Liu, Shizhu He, and Jun Zhao. 2016. How to generate a good word embedding. IEEE Intelligent Systems 31, 6 (2016), 5–14.
[12] David Meyer. 2016. How exactly does word2vec work? (2016).
[13] Sidharth Mudgal, Han Li, Theodoros Rekatsinas, AnHai Doan, Youngchoon Park, and David Meyer. 2016. How exactly does word2vec work? (2016).
[14] Maryam M Najafabadi, Flavio Villanustre, Taghi M Khoshgoftaar, Naeem Seliya, Maxim Mozgovoy, and Vitaly Klyuev. 2014. Automatic detection of antisocial behaviour in texts. Informatica 38, 1 (2014).
[15] Myriam Munezero, Calkin Suero Montero, Tzuom Kakkonen, Erkki Sutinen, and Harith Alani. 2015. Deep learning applications and challenges in big data analytics. Journal of Big Data 2, 1 (2015), 1.
[16] Sirin将 Abdul Rasool Qureshi, Nasirullah Memon, Uffe Kock Will, and Panagiotis Karampelas. 2011. Detecting social polarization and radicalization. International Journal of Machine Learning and Computing 1, 1 (2011), 49.
[17] Maciej Rybinski, William Müller, Javier Del Ser, Miren Nekane Bilbao, and José F Aldana-Montes. 2018. On the Design and Tuning of Machine Learning Models for Language Toxicity Classification in Online Platforms. In International Symposium on Intelligent and Distributed Computing. Springer, 329–343.
[18] Antonela Tommasel, Juan Manuel Rodriguez, and Daniela Godoy. 2018. Textual aggression detection through deep learning. In Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018). 177–187.
[20] Feng Xu, Duoyong Sun, Ze Li, and Bo Li. 2017. Research on online supporting community of extreme organization by AI-SNA based method. In 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS). IEEE, 546–551.

[21] Chunting Zhou, Chonglin Sun, Zhiyuan Liu, and Francis Lau. 2015. A C-LSTM neural network for text classification. arXiv preprint arXiv:1511.08630 (2015).

[22] Peng Zhou, Zhenyu Qi, Suncong Zheng, Jiaming Xu, Hongyun Bao, and Bo Xu. 2016. Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling. arXiv preprint arXiv:1611.06639 (2016).