COLING 2018

First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018)

Proceedings of the Workshop

August 25, 2018
Santa Fe, New Mexico, USA

ISBN 978-1-948087-60-5

Introduction

In the last few years, we have witnessed a gradual shift from largely static, read-only web to quickly expanding user-generated web. There has been an exponential growth in the availability and use of online platforms where users can post their own content. A major part of these platforms include social media websites and apps, blogs, Q&A forums and several similar platforms. All of these are almost exclusively user-generated websites. In all of these platforms and forums, humongous amount of data is created and circulated every minute. It has been estimated that there has been an increase of approximately 25% in the number of tweets per minutes and 22% increase in the number of Facebook posts per minute in the last 3 years. It is posited that approximately 500 million tweets are sent per day, 4.3 billion Facebook messages are posted and more than 200 million emails are sent each day, and approximately 2 million new blog posts are created daily over the web [1]. There is no such a thing as a 'consolidated figure' of the number of comments and opinion generated on websites worldwide, but it can be safely assumed that such a figure would be staggering.

As the number of people and this interaction over the web has increased, incidents of aggression and related activities like trolling, cyberbullying, flaming, hate speech, etc. have also increased manifold across the globe. The reach and extent of Internet has given such incidents unprecedented power and influence to affect the lives of billions of people. It has been reported that such incidents of online aggression and abuse have not only created mental and psychological health issues for web users, but they have in fact forced many people to change things in their daily lives, spanning deactivating accounts to instances of self-harm and suicide. Thus, incidents of online aggressive behaviour have become a major source of social conflict, with a potential of forming criminal activity. Therefore, it is a timely task for researchers and stakeholders to create preventive measures to safeguard the interests of web users, and to contribute to the maintenance of the civility of the online space in a more general sense.

This workshop focusses on the phenomena of online aggression, trolling, cyberbullying and other related phenomena, in both text (especially social media) and speech. The organisers aim to create a platform for academic discussions on this phenomena, based on previous joint work that they have done as part of a project funded by the British Council. We are particularly interested in promoting conversations dedicated to the automatic detection of aggression in both speech and text, that is, we hope that our workshop will not only be purely academic by nature but it will also generate real-life solutions to tackle the phenomena studied. As such the workshop also includes a shared task on 'Aggression Identification'. The task is to develop a classifier that could make a 3-way classification in between 'Overtly Aggressive', 'Covertly Aggressive' and 'Non-aggressive' text data. We made available a dataset of 15,000 aggression-annotated Facebook Posts and Comments each in Hindi (in both Roman and Devanagari script) and English for training and validation. Additional data for testing was released at a later date.

Both the workshop and the shared task received a very encouraging response from the community. There were more than 130 registrations for the shared task. Out of these, 30 teams submitted their systems. The proceedings include 18 system description papers that were finally submitted by the authors. In addition to this, the workshop also includes 5 regular papers presented in the workshop.

We would like to thank all the authors for their submision and members of the Program Committee for their invaluable efforts in reviewing and providing feedback to all the papers. We would also like to thank all the members of the Organising Comittee who have helped immensely in various aspects of the organisation of the workshop and the shared task.

[1] Source: https://www.gwava.com/blog/internet-data-created-daily/

Workshops Chairs

Ritesh Kumar, Dr. Bhimrao Ambedkar University, India
Daniel Kadar, Research Institute for Linguistics, Hungarian Academy of Sciences, Hungary

Organising Committee

Ritesh Kumar, Dr. Bhimrao Ambedkar University, India
Daniel Kadar, Research Institute for Linguistics, Hungarian Academy of Sciences, Hungary
Atul Kr. Ojha, Jawaharlal Nehru University, India
Bornini Lahiri, Jadavpur University, India
Marcos Zampieri, University of Wolverhampton, United Kingdom
Mayank, Jawaharlal Nehru University, India
Shervin Malmasi, Harvard Medical School, United States
Abdul Basit, Dr. Bhimrao Ambedkar University, India
Deepak Alok, Rutgers University, United States

Shared Task Organising Committee

Ritesh Kumar, Dr. Bhimrao Ambedkar University, India
Atul Kr. Ojha, Jawaharlal Nehru University, India
Marcos Zampieri, University of Wolverhampton, United Kingdom
Shervin Malmasi, Harvard Medical School, United States

Editors

Ritesh Kumar, Dr. Bhimrao Ambedkar University, India
Atul Kr. Ojha, Jawaharlal Nehru University, India
Marcos Zampieri, University of Wolverhampton, United Kingdom
Shervin Malmasi, Harvard Medical School, United States

Programme Committee

A. Seza Doğruöz, Tilburg University, Netherlands
Adrián Pastor López Monroy, University of Houston, USA
Amitava Das, IIIT-Sri City, India
Asif Ekbal, IIT-Patna, India
Atul Kr. Ojha, JNU, New Delhi, India
Bruno Emanuel Martins, University of Lisbon, Portugal
Cheng-Te Li, National Cheng Kung University, Taiwan
Chuan-Jie Lin, National Taiwan Ocean University, Taiwan
Claudia Peersman, Lancaster University, UK
Cynthia van Hee, LT3, Ghent University, Belgium
Danilo Croce, University of Roma, Italy
Dennis Tenen, Columbia University, USA
Elizabeth Losh, William and Mary College, USA
Els Lefever, LT3, Ghent University, Belgium
Erik Velldal, University of Oslo, Norway
Eshwar Chandresekharan, Georgia Tech, USA
Fumito Masui, Kitami Institute of Technology, Japan
Girish Nath Jha, Jawaharlal Nehru University, India
Haris Papageorgiou, ATHENA Research and Innovation Center, Greece
Hugo Jair Escalante, INAOE, Mexico
Ingmar Weber, Qatar Computing Research Institute, Qatar
Jen Golbeck, University of Maryland, USA
Jacqueline Wernimont, Arizona State University, USA
Kalika Bali, MSRI Bangalore, India
Lee Gillam, University of Surrey, UK
Liang-Chih Yu, Yuan Ze University, Taiwan
Libby Hemphill, University of Michigan, USA
Lun-Wei Ku, Academia Sinica, Taiwan
Mainack Mondal, University of Chicago, USA
Manuel Montes-y-Gómez, INAOE, Mexico
Marco Guerini, Fondazione Bruno Kessler, Trento
Marcos Zampieri, University of Wolverhampton, UK
Matthew Fuller, University of London, UK
Michael Wiegand, Saarland University, Germany
Michael Paul, University of Colorado Boulder, USA
Min-Yuh Day, Tamkong University, Taiwan
Ming-Feng Tsai, National Chengchi University, Taiwan
Monojit Choudhury, MSRI Bangalore, India
Michal Ptaszynski, Kitami Institute of Technology, Japan
Nemanja Djuric, Uber ATC, USA
Pawan Goyal, IIT-Kharagpur, India
Pete Burnap, Cardiff University, UK
Preslav Nakov, Qatar Computing Research Institute, Qatar
Ritesh Kumar, Dr. B. R. Ambedakar University, India
Roman Klinger, University of Stuttgart, Germany
Ruifeng Xu, Harbin Institute of Technology, China
Saptarshi Ghosh, IIT-Kharagpur, India
Sara E. Garza, Universidad Autónoma de Nuevo León (UANL), Mexico

Shervin Malmasi, Harvard Medical School, USA
Thamar Solorio, University of Houston, USA
Thiago Galery, DBpedia Association and Idio, London, UK
Veronique Hoste, LT3, Ghent University, Belgium
Vladan Radosavljevic, Temple University, USA
William Wang, University of California-Santa Barbara, USA
Xavier Tannier, Université Paris-Sud, LIMSI, CNRS, France
Ye Tian, Amazon Cambridge Development Centre (Alexa), UK
Yelena Mejova, Qatar Computing Research Institute, Qatar
Zeerak Waseem, University of Sheffield, UK
Zhunchen Luo, China Defense Science and Technology Information Center, China

Table of Contents

Conference Program

Saturday August 25, 2018

9:00–10:30 Inaugural Session

9:00–9:10 ***Welcome by Workshop Chairs***

9:10–9:30 *Benchmarking Aggression Identification in Social Media*
Ritesh Kumar, Atul Kr. Ojha, Shervin Malmasi and Marcos Zampieri

9:30–10:30 *Keynote Talk*
Rada Mihalcea, University of Michigan, USA

10:30–11:00 ***Coffee Break***

11:00–12:30 Poster Session

RiTUAL-UH at TRAC 2018 Shared Task: Aggression Identification
Niloofar Safi Samghabadi, Deepthi Mave, Sudipta Kar and Thamar Solorio

IRIT at TRAC 2018
Faneva Ramiandrisoa and Josiane Mothe

Fully Connected Neural Network with Advance Preprocessor to Identify Aggression over Facebook and Twitter
Kashyap Raiyani, Teresa Gonçalves, Paulo Quaresma and Vitor Beires Nogueira

Cyberbullying Intervention Based on Convolutional Neural Networks
Qianjia Huang, Diana Inkpen, Jianhong Zhang and David Van Bruwaene

LSTMs with Attention for Aggression Detection
Nishant Nikhil, Ramit Pahwa, Mehul Kumar Nirala and Rohan Khilnani

TRAC-1 Shared Task on Aggression Identification: IIT(ISM)@COLING'18
Ritesh Kumar, Guggilla Bhanodai, Rajendra Pamula and Maheshwar Reddy Chennuru

An Ensemble Approach for Aggression Identification in English and Hindi Text
Arjun Roy, Prashant Kapil, KINGSHUK BASAK and Asif Ekbal

Aggression Identification and Multi Lingual Word Embeddings
Thiago Galery and Efstathios Charitos

A K-Competitive Autoencoder for Aggression Detection in Social Media Text
Promita Maitra and Ritesh Sarkhel

Aggression Detection in Social Media: Using Deep Neural Networks, Data Augmentation, and Pseudo Labeling
Segun Taofeek Aroyehun and Alexander Gelbukh

Identifying Aggression and Toxicity in Comments using Capsule Network
Saurabh Srivastava, Prerna Khurana and Vartika Tewari

Degree based Classification of Harmful Speech using Twitter Data
Sanjana Sharma, Saksham Agrawal and Manish Shrivastava

Aggressive Language Identification Using Word Embeddings and Sentiment Features
Constantin Orasan

12:30–13:50 ***Lunch Break***

13:50–15:55 **Paper Session I**

13:50–14:15 *Aggression Detection in Social Media using Deep Neural Networks*
Sreekanth Madisetty and Maunendra Sankar Desarkar

14:15–14:40 *Merging Datasets for Aggressive Text Identification*
Paula Fortuna, José Ferreira, Luiz Pires, Guilherme Routar and Sérgio Nunes

14:40–15:05 *Cyberbullying Detection Task: the EBSI-LIA-UNAM System (ELU) at COLING'18 TRAC-1*
Ignacio Arroyo-Fernández, Dominic Forest, Juan-Manuel Torres-Moreno, Mauricio Carrasco-Ruiz, Thomas Legeleux and Karen Joannette

15:05–15:30 *Aggression Identification Using Deep Learning and Data Augmentation*
Julian Risch and Ralf Krestel

15:30–15:55 *Cyber-aggression Detection using Cross Segment-and-Concatenate Multi-Task Learning from Text*
Ahmed Husseini Orabi, Mahmoud Husseini Orabi, Qianjia Huang, Diana Inkpen and David Van Bruwaene

15:55–16:20 ***Coffee Break***

Saturday August 25, 2018 (continued)

16:20–18:00 Paper Session II

16:20–16:45 *Delete or not Delete? Semi-Automatic Comment Moderation for the Newsroom*
Julian Risch and Ralf Krestel

16:45–17:10 *Textual Aggression Detection through Deep Learning*
Antonela Tommasel, Juan Manuel Rodriguez and Daniela Godoy

17:10–17:35 *Combining Shallow and Deep Learning for Aggressive Text Detection*
Viktor Golem, Mladen Karan and Jan Šnajder

17:35–18:00 *Filtering Aggression from the Multilingual Social Media Feed*
sandip modha, Prasenjit Majumder and Thomas Mandl

18:00–18:10 Closing