Neurocomputing 500 (2022) 241–242

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Preface

Editorial: Special issue on "Learning to combat online hostile posts in regional languages during emergency situations"

ARTICLE INFO

Article history:
Available online 18 May 2022

Keywords:
Hostile posts
Fake news
Regional languages

ABSTRACT

The current special issue of Neurocomputing was designed to encourage researchers from interdisciplinary domains working on multilingual social media analytics to think beyond the conventional way of combating online hostile posts. The special issue was primarily based on the theme of the First Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation (CONSTRAINT) (https://lcs2.iiitd.edu.in/CONSTRAINT-2021/), held with AAAI'2021. We invited a few good quality papers accepted in CONSTRAINT to submit an extended version. We also made the call open for the general audience. The special issue broadly focused on three major points: (i) **Regional language:** the offensive posts under inspection may be written in low-resource regional languages (e.g., Tamil, Urdu, Bangali, Polish, Czech, Lithuanian, etc.). (ii) **Emergency situation:** The proposed solutions should be able to tackle misinformation during emergency situations where, due to the lack of enough historical data, learning models need to adopt additional intelligence to handle emerging and novel posts. (iii) **Early detection:** Since the impact of misinformation during emergency situations is highly detrimental to the society (e.g., health-related misadvice during a pandemic can cost human lives), we encourage the solutions to be able to detect such hostile posts as early as possible after their appearance on social media.

The accessibility of the Internet has dramatically changed the way we consume information. The ease of use of social media not only encouraged individuals to freely express their opinion (freedom of speech) and provided content polluters with ecosystems to spread hostile posts (e.g., hate speech, fake news, cyberbullying, propaganda, etc.). Such hostile activities increase manifold during political elections and more recently, during the ongoing COVID-19 pandemic. Notably, most hostile posts are written in regional languages, and therefore can easily evade online surveillance engines that typically target resource-rich languages such as English and Chinese. Therefore, regions such as Asia, Africa, and South America, where low-resource regional languages are used for day-to-day communication, suffer due to the lack of tools, benchmark datasets and learning techniques. Other countries such as Italy and Spain, which have several regional languages (pseudo-low-resource), are not as equipped with sophisticated computational resources as English and are also facing the same issues.

The current special issue of Neurocomputing was designed to encourage researchers from interdisciplinary domains working on multilingual social media analytics to think beyond the conventional way of combating online hostile posts. The special issue was primarily based on the theme of the First Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation (CONSTRAINT),[1] held with AAAI'2021. We invited a few good quality papers accepted in CONSTRAINT to submit an extended version. We also made the call open for general audience. The special issue broadly focused on three major points: (i) **Regional language:** the offensive posts under inspection may be written in low-resource regional languages (e.g., Tamil, Urdu, Bangali, Polish, Czech, Lithuanian, etc.). (ii) **Emergency situation:** The proposed solutions should be able to tackle misinformation during emergency situations where, due to the lack of enough historical data, learning models need to adopt additional intelligence to handle emerging and novel posts. (iii) **Early detection:** Since the impact of misinformation during emergency situations is highly detrimental to the society (e.g., health-related misadvice during a pandemic can cost human lives), we encourage the solutions to be able to detect such hostile posts as early as possible after their appearance on social media.

We accepted a total of six research articles as a part of this special issue. A summary of each of the accepted articles is mentioned below.

The paper, titled "*Knowledge Graph informed Fake News Classification via Heterogeneous Representation Ensembles*" [1], authored by Koloski *et al.*, claimed that the incorporation of knowledge graphs into the text representation learning methods can improve the performance of the fake news detection models by producing the state-of-the-art text representation. According to the authors, this is perhaps the first attempt to evaluate the contribution of knowledge graphs in fake news detection. Due to the heterogeneity of datasets available online, it may not be straightforward to

[1] https://lcs2.iiitd.edu.in/CONSTRAINT-2021/.

https://doi.org/10.1016/j.neucom.2022.05.037
0925-2312/

incorporate such knowledge into the model. The paper showed how one can systematically curate (grounded) subject-predicate-object triplets to construct the knowledge graph and incorporate it into both contextual and non-contextual representation learning methods to learn better document representations. Such knowledge-grounded representation learning methods turned out to be highly effective in fake news detection.

The paper, titled "*A Heuristic-driven Uncertainty based Ensemble Framework for Fake News Detection in Tweets and News Articles*" [2], authored by Dutta *et al.*, proposed a heuristic-driven ensemble framework with an approximate Bayesian neural network that works as a statistical feature fusion network for predictive uncertainty estimation and fake news prediction. The proposed framework was tested on both tweets and news articles for fake news classification. The authors also showed experiments on two large-scale datasets and presented various ablations to show the framework's performance over other methods.

In the paper, titled "*Tackling Cyber-Aggression: Identification and Fine-Grained Categorization of Aggressive Texts on Social Media using Weighted Ensemble of Transformers*" [3], the authors, Sharifa and Hoque, presented a weighted ensemble-based system that can identify and classify Bengali aggressive texts into coarse and fine-grained classes. Scarcity of resources and lack of benchmark Bengali aggressive text dataset are the significant barriers to building a classification system and make it more challenging to implement than in high-resource languages. The authors presented a novel Bengali aggressive text dataset (called 'BAD') with two-level annotations – in the first level, the posts are classified into aggressive and non-aggressive; in the second level, the aggressive posts are categorized into religious, political, verbal and gendered aggression classes. Then they proposed an ensemble approach that combines different versions of the BERT models, namely m-BERT, DistilBERT, Bangla-BERT and XLM-R.

The manuscript, "*Investigating Hostile Post Detection in Hindi*" [4], by Bhatnagar *et al.*, dealt with another low resource language, Hindi, which is the most popular language in India. The authors pointed out the challenges in terms of the language ambiguity, lack of large-scale dataset and scarcity of local context to combat the regional fake news. A new dataset was collected as a part of the experiment. Various Transformer based models along with the multi-task learning frameworks were tested to show the state-of-the-art results.

Sheth *et al.*, in their paper, titled "*Defining and Detecting Toxicity on Social Media: Context and Knowledge are key*" [5], addressed three issues in fake news detection – (i) the social and psychological dimensions of fake news spreaders, (ii) limitations of state-of-the-art fake news detection approaches, and (iii) incorporation of external knowledge in the advanced fake news detection pipeline. The authors proposed a knowledge-infused learning framework that learns different contexts of toxic content from different social perspectives and incorporates them into the detection model. The knowledge infusion was performed at three different levels – shallow, semi-deep and deep infusion.

The paper, titled "*Does Aggression Lead to Hate? Detecting and Reasoning Offensive Traits in Hinglish Code-Mixed Texts*" [6] by Sengupta *et al.*, dealt with a code-mixed language (Hinglish = Hindi + English). The paper started by arguing that most of the existing studies on offensive content detection deal with a limited number of offensive classes mostly due to the lack of sufficient datasets. The authors studied the relationship among five offense traits – aggression, hate, sarcasm, humor, and stance in the Hinglish social media content. They proposed a novel deep unified model to achieve state-of-the-art performance across different traits of offensive content. They further proposed a novel causal importance score to check which abusive keywords contribute more to the offensiveness of social media content.

In summary, the special issue has received significant attention which in turn reflects the need for advanced studies to combat hostile posts, particularly in the regional languages. The contributory papers dealt with both resource-rich languages like English as well as low-resource languages like Bengali, Hindi and Hinglish. Various datasets and novel methods were also proposed. The authors explored psychologically and sociological aspects of hostile posts as well.

We would like to thank Prof. Zidong Wang, the EIC of Neurocomputing and the entire editorial team for their constant support. We also want to thank the authors for their contributions, and the reviewers for their rapid, thorough and timely reviews.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] B. Koloski, T. Stepišnik Perdih, M. Robnik-Šikonja, S. Pollak, B. Škrlj, Knowledge graph informed fake news classification via heterogeneous representation ensembles, Neurocomputing 498 (2022) 208–226.
[2] S.D. Das, A. Basak, S. Dutta, A heuristic-driven uncertainty based ensemble framework for fake news detection in tweets and news articles, Neurocomputing 491 (2022) 607–620.
[3] O. Sharif, M.M. Hoque, Tackling Cyber-Aggression: Identification and Fine-Grained Categorization of Aggressive Texts on Social Media using Weighted Ensemble of Transformers, Neurocomputing 490 (2022) 462–481.
[4] V. Bhatnagar, P. Kumar, P. Bhattacharyya, Investigating Hostile Post Detection in Hindi, Neurocomputing 474 (2022) 60–81.
[5] A. Sheth, V.L. Shalin, U. Kursuncu, Defining and detecting toxicity on social media: context and knowledge are key, Neurocomputing 490 (2022) 312–318.
[6] A. Sengupta, S.K. Bhattacharjee, M.S.a. Akhtar, T. Chakraborty, Does aggression lead to hate? Detecting and reasoning offensive traits in hinglish code-mixed texts, Neurocomputing 488 (2022) 598–617.

Tanmoy Chakraborty [a,*]
Kai Shu [b]
H. Russell Bernard [c]
Huan Liu [c]
[a] *IIIT-Delhi, India*
[b] *Illinois Institute of Technology, USA*
[c] *Arizona State University, USA*
* Corresponding author.
E-mail addresses: tanmoy@iiitd.ac.in (T. Chakraborty), kshu@iit.edu (K. Shu), asuruss@asu.edu (H.R. Bernard), huanliu@asu.edu (H. Liu)
Received 8 May 2022
Accepted 14 May 2022

Available online 18 May 2022