Identifying Complaint Based Tweets in Social Media Platforms:A Review

1st Naman Goenka
Department of Computer Science and Information Systems
Birla Institute of Technology and Science, Pilani
Rajasthan, India
f20180398@pilani.bits-pilani.ac.in

***Abstract*—This review paper has been written in partial fulfilment of the selection process for intern role in MIDAS, IIIT Delhi. We attempt to review and present our understanding of possible improvements, future scope and shortcomings of one research paper related to the topic of Identifying Complaint Based Tweets in Social Media Platforms. It commences with an brief introduction of the approach taken, then proceeds with scope of future work of the topic and improvements possible and finally concludes with shortcomings identified in the approach followed.**

***Index Terms*—informative cues, domain relevance score, seed phrases, transport domain**

I. Introduction and Future Scope of the work

Authors presented a novel semi-supervised pipeline along with a novel dataset for identification of complaint based posts in the transport domain. The proposed method should also be evaluated in in terms of applicability in other domains to evaluate out generalisability of the approach and bias towards the domain of the experiment. Moreover additional features such as demographic features, contextual features, network information features, features denoting virality of the tweets, etc. Demographic features being the most important to separate high noise in obtained data versus data points being influential in accommodating demographic variability.

Additionally, we can also replace use of tf-idf representations throughout our study with more recent state-of-the-art techniques. Data augmentation and knowledge transfer techniques can also be applied enabling use of neural techniques on the task at hand, enabling us to learn more complex features varying with different types of complaints.

Metadata for the tweets can also be used as features for evaluating whether there is any relation between retweet pattern, interaction, the time of tweet and user background, etc, in the tweets being complaint relevant.

II. Shortcomings identified in approach followed

Analysing the best performing model,Intensifiers, in information specific features which has second best performance overall might indicate that the approach was able to identify complaint based tweets with information cues perhaps based on intensity in expression as a key semantic feature. But, a whole subset of transport based complaint related to a less intensified, positive assertive tone might be completely left out. They will perhaps be more valuable to the organisation in terms of providing valuable suggestions or feedback from users, associated with a complaint.

The part of the experiment in which, a forum was used for providing initial lexicons for further extraction of tweets should be placed with more robust way to do so such that extracted lexicons incorporate variability in terms of demography; type of service upon which complaint is based, extent of difficulty faced by the user; seriousness of the issue ,for example, a crime is being reported; proportion of actual complaints being posted on the forum, etc. One solution that I would like to propose is to ensemble posts from different forums,preferably including government platforms too, and then using it for lexical analysis.

References

[1] Anand, Gyanesh, Akash Kumar Gautam, Puneet Mathur, Debanjan Mahata, Rajiv Ratn Shah, and Ramit Sawhney. "An Iterative Approach for Identifying Complaint Based Tweets in Social Media Platforms (Student Abstract)." In AAAI, pp. 13749-13750. 2020.