

Towards Selecting the Right Social Media Platform Based on Journalistic Requirement

Shalini Priya, Ryan Sequeira, Joydeep Chandra, Sourav Kumar Dandapat

Deptment of Computer Science and Engineering

Indian Institute of Technology Patna

Patna, India-801103

Email: {shalini.pcs16, ryan.mtcs16, joydeep, sourav}@iitp.ac.in

Abstract—Social media got huge popularity in last decade. People use this media for reporting the event and sharing their opinion. It has been found that in numerous cases events are being first reported in social media. Hence, social media platforms have become a natural choice for news agencies for news searching, news aggregation, mining of opinions and to get feedback from readers. However, all social media are not equally informative and hence not suitable for all kinds of journalistic requirement. This study explores that there is significant differences across these platforms that can be suitably exploited depending upon the scope of the requirements.

I. NON-TRIVIALITY OF THE PROBLEM

Characteristics of Twitter and Reddit in terms of social structure (in Twitter user follows another user where in Reddit user subscribes to subreddits), posting pattern (restriction in post length: users on twitter are restricted to write a Tweet within 140 characters where there is no bound in Reddit, posting frequency), user-base (for reddit, users subscribed to a subreddit is having some expertise with high probability) are very different. Head-to-head comparison of two such media is inherently challenging.

II. PROPOSED APPROACH

We have identified the news categories from a set of 500 events that were widely discussed on both the platforms Twitter and Reddit. The flow of our current research can easily be understood using Figure 1.

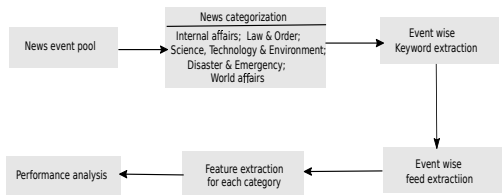


Fig. 1: Block Diagram of Proposed Approach

We have studied the characteristics difference of Twitter and Reddit in terms of content features, temporal features and social features for different categories of news events. Content features include information richness, opinion diversity, Average number of slangs and non-dictionary words. Temporal features include spreading pattern, Lifespan of the event, Inter-arrival time of comments/tweets. Social features include a mean number of influential users, Up-votes and re-tweet count, the Mean activity rate of the users of Reddit and Twitter.

III. CURRENT STATUS AND PRELIMINARY RESULTS

Figure 2 compares the information richness of comment/tweet for each of the news category. It is calculated with the help of additional (which are not present in the related article) named entity and noun phrase present in comment/tweet. As can be observed, the mean additional information in Reddit comments is much higher for all the news categories, indicating that Reddit can provide more additional relevant information as compared to Twitter, and it can be helpful in writing the blogs. It is also evident from the figure that the mean additional information increases with the increase of comment length. Events related to disaster and emergency has minimum comment length and it provides minimum additional information. The burstiness of the comments and feeds can be analyzed more precisely from figure 3. Figure shows that irrespective of news category, first 50% of Reddit comments are received within first 10% time while 50% of feed arrived in first 40% of time in Twitter. This implies that, we can expect a heavy burstiness in Reddit comments in the first phase of event and within very short time we can estimate lot of characteristics of an event.

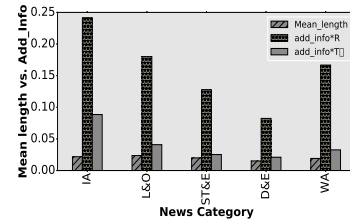


Fig. 2: Mean Additional Information

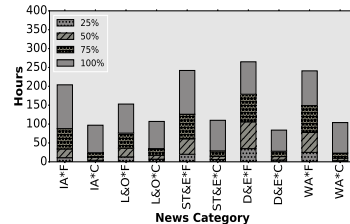


Fig. 3: Burstiness

IV. FUTURE WORK

Our future work is to find more platform, features and scrutinize their characteristic on both the platform and suggest suitable OSN based on journalistic requirement.