Acta Scientific AGRICULTURE (ISSN: 2581-365X)

Volume 5 Issue 5 May 2021

Editorial

Information Filtering and Organisation in Social Microblogging Sites

Soumi Dutta*

Department of Computer Science and Applications, Institute of Engineering and Management, India

***Corresponding Author:** Soumi Dutta, Department of Computer Science and Applications, Institute of Engineering and Management, India.

Received: March 22, 2021

Published: March 29, 2021

Online social microblogging sites, such as Twitter (https://twitter.com), Tumblr (http://www.tumblr.com), and Sina Weibo (https://weibo.com) are very popular platforms for information exchange on the Web today. These micro blogging sites have become trendy communication tools, enabling rapid information exchange. Vast amounts of information are generated on these sites every day through economic, academic and social activities. This crowd-sourced information can be utilised for diverse applications, ranging from market analysis, fraud detection, detection of spammers and spam posts, characterization or grouping of user according to their behavior, extraction of important news and customer retention to production control and science exploration. Data extracted from microblogs are increasingly being used to build real-time search and recommender systems, and services which mine and summarize public reactions to events. Side by side with the diverse applications, microblogging sites also bring in several challenges in utilizing the crowd-sourced data, notably need for filtering out malicious content posted by spammers and need for organizing the plethora of information.

- **Challenges in using microblogging sites:** Social microblogging sites are increasingly being used as information media, to get up-to-date information on various topics of interest, including various ongoing events such as socio-political events, sports events, natural disasters, etc. Hence methodologies need to be developed for helping users utilize such information systems. The thesis focuses primarily on two challenges that need to be addressed to this end.
- **Filtering harmful information:** There is a wide variety in the quality of information posted on online social media, ranging from important real-time information on various topics posted by authoritative sources, to various types of malicious content including spam, hate speech, etc. Especially spam posts are a severe problem in media like Twitter. Hence, one of the problems is to accurately identify spam posts, and users who post such spam, in real-time.
- **Organizing the information**: A huge amount of information is regularly posted on social microblogging sites. For instance, over 500 million tweets are posted per day on Twitter. Even from the point of view of an individual user, a reasonably active Twitter user can easily get hundreds of tweets in his/her timeline every day. This plethora of content leads to information overload, and no user can digest so much information. So it is another challenge to address the information overload problem. Hence methodologies for organizing the information are required.

There are already several algorithms for text summarization, most of which are extractive in nature. It is observed that different summarization algorithms generate very different summaries from the same input set of microblogs. Rather than trying to develop a new summarization algorithm, the ensemble algorithms can be proposed that attempt to combine the outputs of multiple base summarization algorithms, to produce a summary that is better than what is obtained from any of the individual base algorithms.