

Resolving the where and when of social media events

September 24 2021

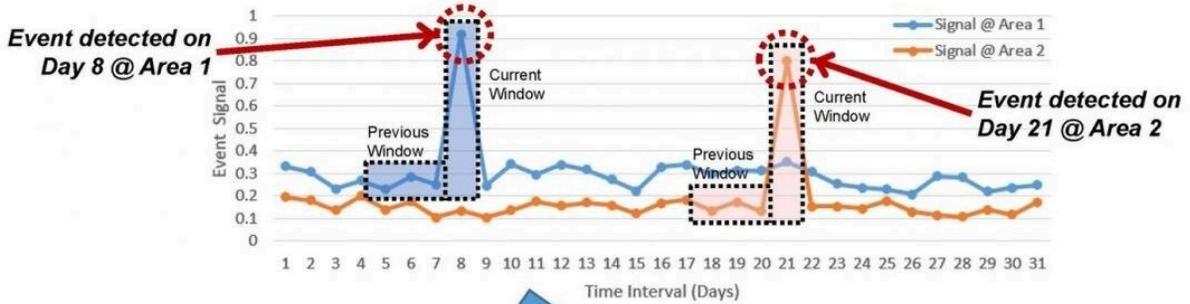
Phase 0: Retrieve Geo-tagged Tweets



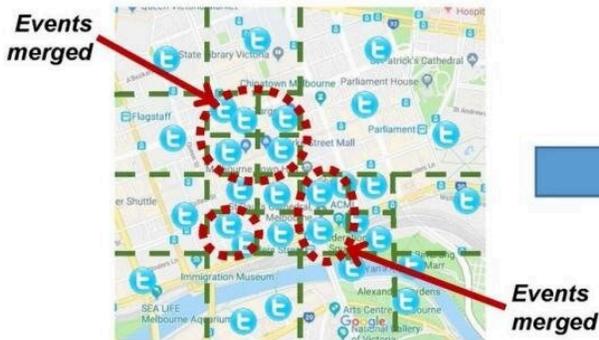
Phase 1: Build Quad-tree



Phase 2: Detect Events



Phase 3: Merge Events



Phase 4: Prune Events



Overview of our proposed Spatio-temporal Online Event Detection Algorithm.
Credit: SUTD and MoU

Researchers from the University of Melbourne (UoM) and the Singapore University of Technology and Design (SUTD) have developed an algorithm that can detect important events based on the time and geographical scale of topics being actively discussed on social media. Their algorithm, detailed in the *Journal of Big Data*, does not require knowing which events to detect upfront and can be tailored to use smaller or larger geographical and time resolutions to reflect the dynamic nature of real-life events.

Social media has become the go-to medium for communication, largely because posts can be uploaded and disseminated almost instantly. Given that many users actively share observations and photos of events happening around them, such real-time information makes [social media](#) an attractive source of breaking news.

However, with more than two-thirds of people on the internet—or 2.5 billion people—using these platforms globally, strategies are needed to wade through the noise to extract useful event-specific data. For life-threatening occurrences requiring emergency and security personnel, the need for immediate, event-specific information is all the more acute.

"Elements of time and space give you a better resolution of where and when the events are happening," said study co-author SUTD Assistant Professor Kwan Hui Lim. "If there is any kind of disaster, you want to know where and when it's happening so you can allocate the right resources to that particular location."

The [algorithm](#) developed by Asst Prof Lim and a team led by Professor Shanika Karunasekera from the UoM adopts a four-phase structure to identify events at different space and time resolutions. Given a stream of raw geo-tagged social media posts, the first phase determines the granularity or resolution in space to detect events. This is accomplished by splitting a [geographical area](#) into multiple scales based on the density

of social media posts.

In the second phase, the algorithm uses statistical methods to identify events based on regions with an unexpectedly high or low density of social media activity. These events, which are fixed in time, are merged in the third phase if they occur in the same geographical area at consecutive time intervals—giving an estimated duration for each event. The fourth and final phase then prunes any events that turn out to be noise.

Besides the ability to tailor the scale of space and time for identifying events, the algorithm is unique in that it requires no prior identification of events to be detected. "This has both advantages and disadvantages," Asst Prof Lim noted. "The advantages are that you do not need a pre-existing dataset with pre-labeled events. Without these pre-existing events, you can use the algorithm to detect new events that you have not seen before. The downside is that you have to manually determine the threshold that will trigger an alert from the system."

The researchers validated their algorithm on streams of posts originating from major cities around the world, focusing on two [social media platforms](#) with very different modalities: the microblogging service Twitter and the photo sharing platform Flickr. The new algorithm outperformed two baseline algorithms based on standard metrics such as precision and recall, as well as a novel measure called the strength index developed by the team.

The strength index measures the fraction of top entities, which can include hashtags or mentions on Twitter or image tags and descriptions on Flickr, to the total number of posts about the detected event. According to Asst Prof Lim, other domains that use information retrieval or classification could benefit from applying the strength index.

In the future, the algorithm could be strengthened by making it multimodal, Asst Prof Lim said. Given that people on social media tend to be selective in what they choose to say and vary the type of information they share on different platforms, an algorithm that can combine data from multiple sources, such as from Twitter and Flickr as well as traditional news media, may improve the reliability of event detection.

"Social media is a potential treasure trove of data for governments, news [media](#) and businesses alike. It is particularly useful as a source of breaking news, especially for first responders like emergency and security personnel. This sophisticated algorithm will help these organizations sort through the noise to extract just the necessary information they need," concluded Prof Karunasekera.

More information: Yasmeeen George et al, Real-time spatio-temporal event detection on geotagged social media, *Journal of Big Data* (2021). [DOI: 10.1186/s40537-021-00482-2](https://doi.org/10.1186/s40537-021-00482-2) Yasmeeen George et al, Real-time spatio-temporal event detection on geotagged social media, *Journal of Big Data* (2021). [DOI: 10.1186/s40537-021-00482-2](https://doi.org/10.1186/s40537-021-00482-2)

Provided by Singapore University of Technology and Design

Citation: Resolving the where and when of social media events (2021, September 24) retrieved 19 April 2024 from <https://techxplore.com/news/2021-09-social-media-events.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.