

Narrative Extraction through the Detection and Characterisation of National and Local Events

Nicholas Bennett, David Millard, David Martin

University of Southampton

Extended Abstract

1. Introduction

Social media are popular forms of information dissemination with users potentially accessing an audience numbering in the millions. With the power and popularity of smartphones and other GPS-enabled smart devices, the creation and widespread publication of user-generated location-based content has rapidly grown in recent years. Twitter, a key platform in this trend, has been extensively researched by academics in the past (Crooks et al., 2013; Hamed, 2014); however, the frequency of location information is increasing and the ability to extract it from the text-based data to understand human activity, mobility and events is a key field for academics both from the data-driven and sociological disciplines. Furthermore, mapping companies interested in real-time change and event data would also benefit from a concise understanding of these aspects. The work defines “event” as a union of Pohl et al. (2012), Dou et al. (2012) and Smith et al. (2015)’s definitions: an activity taking place within a physical location that has a social representation. This definition covers such events as music festivals, pop-up markets and other transient geospatial activities.

Existing data-driven event detection and classification methods focus heavily on large-scale events, aiming to extract location information from human reactions to earthquakes, floods and other natural disasters (Sakaki et al., 2010). Recent work has refined these two methods of analysing street-level data (Middleton et al., 2014; Smith et al., 2015); however, the field is still

very new and in much need of further development. With the increase in number of services that monetise location-based understanding, there is a vested interest in researching the area.

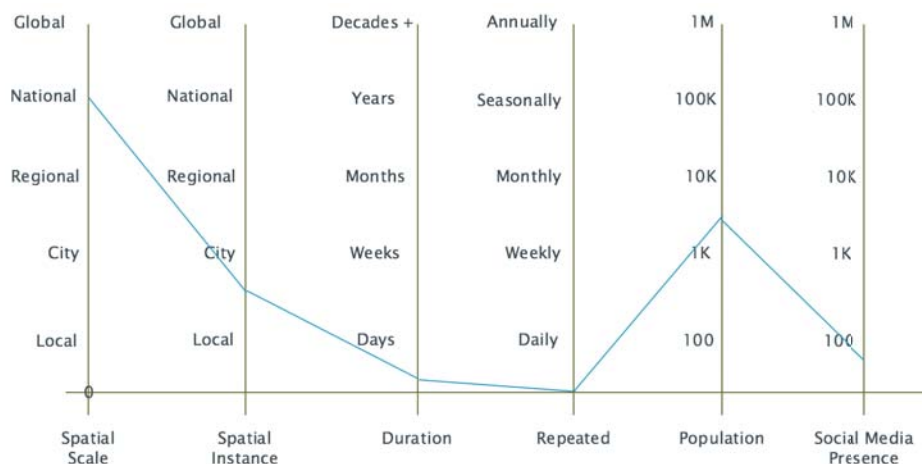


Figure 1: The parallel coordinate graph enables classification of social events. A small music festival example has been plotted.

2. Methodology

Collecting and analysing social media data includes significant linguistic problems, namely discovery of relevant terms and extraction of event descriptions. To mitigate against this, the project captures geolocated tweets from Southampton by applying coordinates to the Twitter API request parameters to obtain over 330,000 tweets from the 9th-20th June 2016. Whilst this method gathers significantly more data than a keyword search, it avoids keyword bias and allows for spontaneous hashtags and peripheral connections to be captured; however, it does include location bias (such as densely populated cities naturally producing more data) but as it is a relatively small boundary box search the results are not limited to Twitter's 1% feed.

A parallel coordinate model is proposed as appropriate for modelling social events (seen in figure 1). The model allows for the representation of several aspects of social events. This permits distinct multivariate profiles for different types of events to aid in their classification, as well as narrowing the scope of applicable events.

To test the model, UK-based event-related BBC News articles are taken as ground truth and applied to the axes. The Twitter data is first analysed for its popular topics and then searched for discussions about the related BBC News events.

3. Preliminary Results

The preliminary results show a discrepancy between nationally important news and topics of interest to the local population. There are often news articles of assumed importance that have no Twitter reflection at all. However, there are certain topics such as the shooting of MP Jo Cox that prominently feature in both the BBC News articles and the Twitter data. Therefore, whilst some news articles are assumed to have national social significance, local interests often differ.

4. Future Work

Only detecting the presence of a social event, however, does not create a holistic understanding. Dynamic sentiment analysis aids in creating a rich, semantic impression of a location. This work aims to extract sentiment for a particular area not only to understand what users are doing but how they are doing it, with whom and how they are feeling at the time. The overlay of narratives onto a physical location adds a new dimension to existing practises. More work is needed to verify these event detection methods such as automatically integrating other social media and web resources to form a consensus, and to incorporate rigorous natural language processing to differentiate between event and non-event related data. Once accurate information is extracted, the Twitter data can be used to plot the event profile without the use of a ground truth article.

The output of this project will be a change and event detection algorithm that pulls social media and web data to create a verified database of semantically rich location information. This methodology will aid in improving existing practises and offer accurate, reliable, real-time change and event information to interested parties.

References

- Crooks A, Croitoru A, Stefanidis A, Radzikowski J (2013) #Earthquake: Twitter as a Distributed Sensor System. *Transactions in GIS*, 17(1), 124–147.
- Dou W, Wang X, Ribarsky W, Zhou M (2012) Event Detection in Social Media Data. IBM Almaden Research Center, (pp. 2–5).
- Hamed A (2014). An Exploratory Analysis of Twitter Keyword-Hashtag Networks and Knowledge Discovery Applications. Doctoral dissertation, University of Vermont.
- Middleton S E, Middleton L, Modafferi S (2014) Real-time crisis mapping of natural disasters using social media. *IEEE Intelligent Systems*, 29(2), 9–17.
- Pohl D, Bouchachia A, Hellwagner H (2012) Automatic sub-event detection in emergency management using social media. *Proceedings of the 21st international conference companion on World Wide Web - WWW '12 Companion*, (pp. 683).
- Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes Twitter users: real-time event detection by social sensors. *WWW '10: Proceedings of the 19th international conference on World wide web*, (pp. 851).
- Smith L, Liang Q, James P, Lin W (2015) Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework. *Journal of Flood Risk Management*.