

Week 7-2: Paper Summaries

CE-510 Seminar: Social Media Mining

Student Name: Jiaqi Guo NetID: JGR9647

■ Twitter location (sometimes) matters: Exploring the relationship between georeferenced tweet content and nearby feature classes

In this paper, the author investigates whether microblogging texts (tweets) produced on mobile devices are related to the geographical locations where they were posted. To do this, they associated Twitter topics with regions. In this process, OpenStreetMap's classification points of interest can serve as verification points. In term of methodology, they use manual, supervised, and unsupervised machine learning methods to classify and locate these points, which are associated with tweet content.

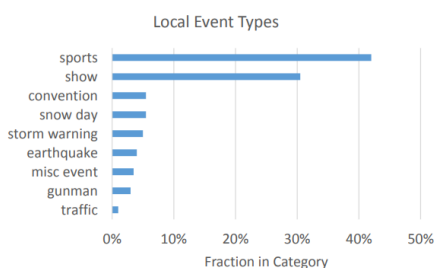
Possible Improvement Directions:

1. The model failed to consider the spatial pattern of tweets, the scale of analyses of tweets needs to be adapted to their spatial resolution.
2. The result of experiments has shown that treating tweets as being relevant to a set of coordinates with precision of the order of tens of meters is unlikely to be a sensible approach to exploring such data. There is a pressing need to more critically consider the extent to which the coordinates of a piece of information can be related to location by considering issues such as scale, abstraction and more cognitively adequate tessellations of space.

■ Eyewitness: Identifying Local Events via Space-Time Signals in Twitter Feeds

In this paper, the researchers present a methodology for automatically extracting and summarizing reports of significant local events from large-scale Twitter feeds. The previous work has largely relied on an analysis of the tweet text to identify local event, they use only time series analysis of geotagged tweet volumes from localized regions. In terms of methodologies, their proposed event detector works by finding where the number of actual tweets exceeds the number of predicted tweets by three times the standard deviation of the prediction error.

Possible Improvement Directions:



1. The type of topic is limited, and events are more likely to be found if they contain proper nouns. It is not difficult to find from this table that "sports" and "show" occupy a large proportion, which may be because sensitive keywords such as team name, name and location are included in such events. The future development direction can be to find a way to make the detector more sensitive to some general events.

2. The occurrence range of events may be different, for example, the impact range of a traffic jam and an earthquake may be significantly different. Therefore, how to choose the observation resolution of detector is a problem that can be studied