Geospatial Event Detection through Tweets

Yuqian Huang, Claudia E. Aviles Toledo, Yue Li, Jie Shan (jshan@purdue.edu)
Lyles School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA

Introduction

Objectives

Through the use of social media, many crowd-sourced sensing and collaboration projects can benefit from vast amount **Volunteered Geographic Information** (VGI). And this study is trying to find out **spatial and temporal pattern** of individual users as well as potential **events** and other interesting findings.

Data

The data for this study is the geo-tagged Tweets between 2014 and 2015 for four college cities: **Purdue, West Lafayette; The Ohio State University, Columbus; Indiana University, Bloomington; and University of Michigan, Ann Arbor**. There are totally **5,230,757** tweets in these four cities.

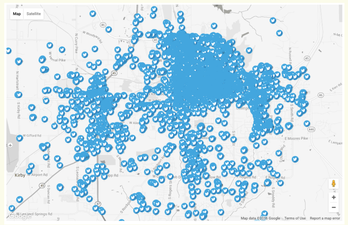


Fig. 1 Tweet Samples at Bloomington in Apr 2014

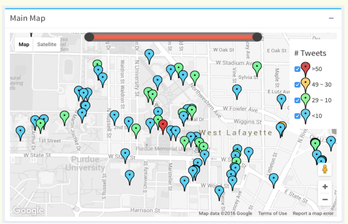


Fig. 2 Tweets samples from an event around Purdue Campus in Dec 2014

Methodology

- Clustering: Density-Based Spatial Clustering of Applications with Noise (**DBSCAN**) for individual pattern, and **ST-DBSCAN** (**ST:** spatial and temporal), a density-based clustering algorithm based on DBSCAN, for event detection
- Keyword Detection: Descending order of word frequency
- Transferring attributes between different Layers: Spatial Joint

Pattern of an Individual

1. Group individual tweets hourly and apply DBSCAN to detect potential clusters.

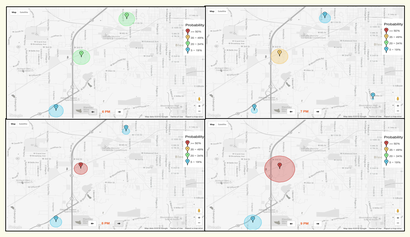


Fig. 3 Pattern of a Twitter user in BL in 6-7, 7-8, 8-9,9-10PM for Year 2014

2. Summarize the contribution of different types of clusters into bar chart.

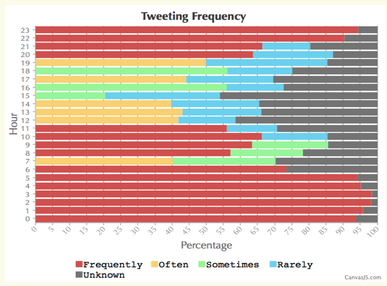


Fig. 4 Temporal Pattern of the user in Fig. 3

Pattern at Road Intersections

This is an example of how Twitter information could help know interest intersections with heavy traffic, crash studies, etc.

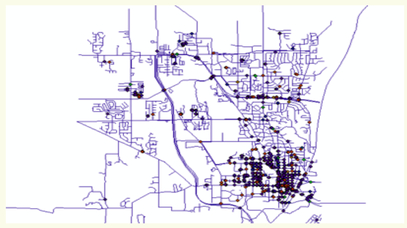


Fig. 5 Spatial Pattern of Tweets at Intersections

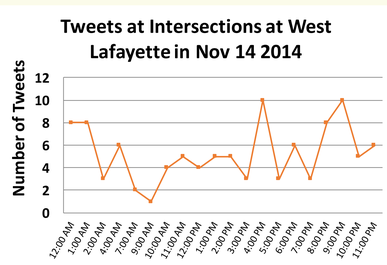


Fig. 6 Temporal Pattern of Tweets at Intersections

Event Study

Event Recall: Gunshot at Purdue

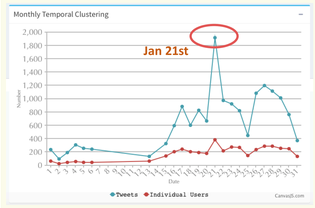


Fig. 7 Monthly Temporal Distribution at West Lafayette in Jan 2014

Description

The shooting happened on the Electrical Engineering building around noon of Jan 21, 2014, and then all students on campus sheltered-in-place. As this was a sudden and shocking event, word spread very quickly and people all over WL, especially students on campus talked about this on Twitter. During the lockdown period, students went on Twitter for latest updates from Purdue official accounts as well as their friends, and they tweeted or retweeted about the event.

Spatial-Temporal Pattern

In Fig. 8, the heat map shows several hot spots: Spot 1 is Purdue Memorial Union, the nearest hot spot to the crime scene, while 2, 3, 4 are student dorms. For the temporal part, the outbreak of tweets happened 12:30 PM - 01:30 PM, which is short after the event, and it slowly faded out after 3:30 PM.

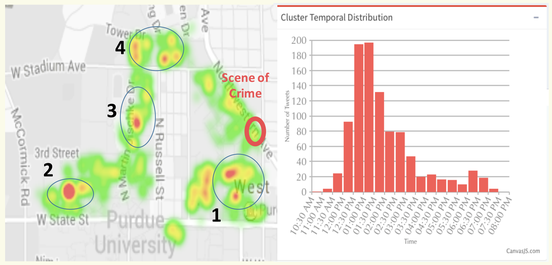


Fig. 8 Heat Map (Left) and Temporal Distribution (Right) of the Gunshot Clusters

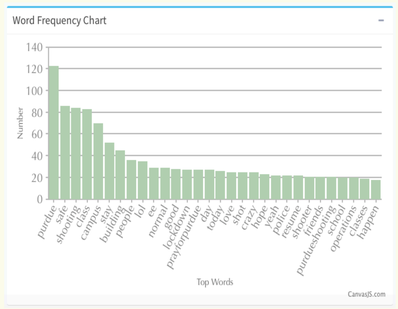


Fig. 9 Frequency of Top 30 Words in the Tweet Clusters

Textual Information

'purdue' turned out to be hottest word. It likely meant some event highly related to Purdue, since most Twitter users on campus usually don't use it. The words like 'shooting', 'shot', 'shooter', 'police', 'purdueshooting' show that the gunshot event spread among the whole campus. The words like 'prayforpurdue', 'love', 'crazy', 'hope' reflect some emotions and attitude from Twitter users towards this event. The word 'ee' stands for Electrical Engineering Building which is the scene of crime.

Event Detected: Beer Festival in Bloomington

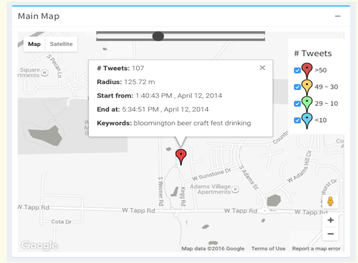


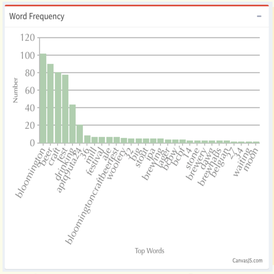
Fig. 10 Information of the found Cluster on Apr 12th 2014 in Bloomington

Description

Daily spatial temporal clustering is applied. Find the centers with # tweets more than a threshold (e.g.50). We did not have any previous knowledge of this event, but we found a cluster to the southwest of Bloomington Downtown on Apr 12th 2014. This cluster has some very interesting keywords: beer, craft, fest, and drinking, which highly indicates that there might be a beer-drinking event.

Textual Information

'bloomington', 'beer', 'craft', 'fest', and 'drinking' dominated the word frequency chart, while other words are with low frequency. The most common pattern of the text is "Drinking ***** @ bloomington craft beer fest", which is the major reason of the structure of word frequency distribution chart.



- Bloomington Brew Fest ! (@ Woolery Mill w/ 7 others)
- About to begin! #craftbeerfestival @ Woolery Mill
- Beer o'clock! @ Bloomington Craft Beer Festival
- Drinking a War Mullet by @ bloomington craft beer fest
- Drinking a Zombie Dust (2014) by @ bloomington craft beer fest
- Drinking a Rosenweiss by @ bloomington craft beer fest
- Drinking a Star City Lager by @ bloomington craft beer fest

Fig.11 Frequency of Top 30 Words in the Cluster (Left) and Sample Texts from Tweets (Right)

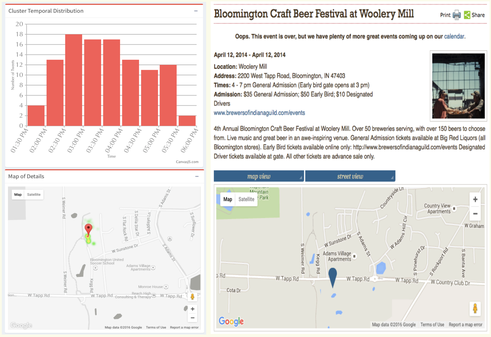


Fig.12 Detected Cluster's Temporal Pattern (Tope Left) , Spatial Pattern (Bottom Left), and the posted Beer Festival Information (Right)

Event Conformation

- For the temporal pattern, the posted period is 3 – 7 pm, while the detected one is 1:30 – 6 pm. Likely some people were calling their friends or making reservations online before the festival, and the people might leave earlier.
- For spatial pattern, the event was at Woolery Millis, which coincided with the detected cluster. Note the posted event address, shown in the middle of a road (blue dot), was wrong.

Conclusion

This project successfully finds out and describes spatial and temporal pattern of individual users. We take advantages of Twitter data to recall and detect events as well as to analyze their spatial-temporal patterns and textual information, and to assess the importance of events.

This would potentially help emergency planning, risk management and damage assessment. There are still improvements to make, and one of which is topic classification. A clear understanding of potential different topics in one cluster could discover multiple events in one cluster.

References

Huang, Qunying, and David WS Wong. "Modeling and visualizing regular human mobility patterns with uncertainty: an example using Twitter data." *Annals of the Association of American Geographers* 105.6 (2015): 1179-1197.
Birant, Derya, and Alp Kut. "ST-DBSCAN: An algorithm for clustering spatial–temporal data." *Data & Knowledge Engineering* 60.1 (2007): 208-221.