Location-Based Social Media Data Mining

Lei Zou, Ph.D.
Postdoctoral Research Associate, Louisiana State University
GIS Workshop @ University of North Texas, Feb-16-2018
Website: www.rsgis.envs.lsu.edu/lzou/gisworkshop.html
Introduction
Introduction

Social Media: Websites or applications that enable users to create and share content or to participate in social networking.

Top 10 Social Media Apps (Source: Dreamgrow, 2018)

Users are using social media in many ways
Introduction

How does social media read my mind?
Introduction
Introduction

Content Layer

Social Layer

Geographical Layer
Outline

1. Introduction to location-based social media data

2. How to conduct location-based social media data mining?

3. A Python demo of Twitter data collection and processing
1. Introduction to location-based social media data

**Definition:** Data collected from social media websites or applications which **provide location-based services** that allow users to “check in” at **physical places**, and directly or indirectly **indicate the location** in the data (modified from Guo and Liu, 2012)
1. Introduction to location-based social media data

Big Data - 5Vs: Volume, Variety, Velocity, Value & Veracity

Social media users in US
2008: 24%
2018: 81%

Twitter: 500 million/day
Facebook: 1 billion/day
Instagram: 40 million/day

Real-time streaming data

Number of Global Social Media Users

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of users in billions</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>0.97</td>
</tr>
<tr>
<td>2011</td>
<td>1.22</td>
</tr>
<tr>
<td>2012</td>
<td>1.4</td>
</tr>
<tr>
<td>2013</td>
<td>1.59</td>
</tr>
<tr>
<td>2014</td>
<td>1.91</td>
</tr>
<tr>
<td>2015</td>
<td>2.14</td>
</tr>
<tr>
<td>2016</td>
<td>2.28</td>
</tr>
<tr>
<td>2017</td>
<td>2.46</td>
</tr>
<tr>
<td>2018*</td>
<td>2.62</td>
</tr>
<tr>
<td>2019*</td>
<td>2.77</td>
</tr>
<tr>
<td>2020*</td>
<td>2.9</td>
</tr>
<tr>
<td>2021*</td>
<td>3.02</td>
</tr>
</tbody>
</table>
1. Introduction to location-based social media data

Geographic Information

Geo-tags
(Place/Coordinates)

User Profile
(Address)

Content
(Associated Location)
1. Introduction to location-based social media data

Users Voluntarily Generated Contents

Events → Users → Contents
1. Introduction to location-based social media data

Applications: Marketing

Step 1: Data Analysis
Preference of models in different regions

Step 2: Advertisement
Distribute different ads to different regions

Facebook Users

Kia America

2.5X More clicks
3.3X More purchase
$13 Lower cost per effective ad
1. Introduction to location-based social media data

Applications: Survey

Facebook fans of NFL teams across the United States based on which official team page has the most 'Likes' from people who live in that county. (The Atlantic, 2014)
1. Introduction to location-based social media data

Applications: Natural Hazard Monitoring

Purpose
Estimate county-level damage by Hurricane Sandy from Twitter data.

Twitter information:
Frequency and content

Results:
Without Twitter: 46%
With Twitter 56%

County-level disaster-related Twitter intensity during Hurricane Sandy
(Zou et al 2018, Annals of the American Association of Geographers.)
1. Introduction to location-based social media data

Applications: Emergency Management

13458  5  37456  780
RESCUERS AND VOLUNTEERS  DISASTERS (SINCE 2017)  SURVIVORS HELPED  CERTIFIED FIRST RESPONDERS

Rescue disaster victims through a social network platform (Crowd Source Rescue, Hurricane Harvey)

http://crowdsourcerescue.com/
Outline

1. Introduction to location-based social media data

2. How to conduct location-based social media data mining?

3. A Python demo of Twitter data collection and processing
2. How to conduct location-based social media data mining?

5-Step Framework

Step 1: Collection

Step 2: Parsing

Step 3: Geocoding

Step 4: Storage

Step 5: Analysis & Visualization
2. How to conduct location-based social media data mining?

5-Step Framework

Step 1: Collection
Step 2: Parsing
Step 3: Geocoding
Step 4: Storage
Step 5: Analysis & Visualization
2. How to conduct location-based social media data mining?

Step 1: Collection

**Web Crawlers**
- Technical challenges
- High risk (account/ip address being blocked)

**Commercial companies**
- Full dataset
- Expensive ($300/one million tweets)

**Application Programming Interface**
- Free for standard use
- ‘Legit’ method to collect data
- Rate limited
2. How to conduct location-based social media data mining?

Step 2: Parsing

The process of analyzing collected social media data, either in natural language, computer languages photos or url links, into useful information.

User: Lei Zou

Sentiment: Positive

Time: 12:04 pm, Feb 9, 2018

Location: Baton Rouge
2. How to conduct location-based social media data mining?

Step 3: Geocoding

Geocoding is the process of converting addresses (like a street address) into geographic coordinates (like latitude and longitude), which you can use to place markers on a map, or position the map.

Input address
University of North Texas

Geocoding
Searching & Matching

Output Coordinates
University of North Texas
(Ing: -97.15, lat: 33.21)

Display
Outline

1. Introduction to location-based social media data

2. How to conduct location-based social media data mining?

3. A Python demo of Twitter data collection and processing
3. A Python demo of Twitter data collection and processing

**Input:**
Keywords (e.g. Dallas)

**Output:**
Heatmap of Tweet Density

**Step 1: Collection**
tweets containing keywords

**Step 2: Parsing**
locations & contents

**Step 3: Geocoding**
determine coordinates

**Step 4: Storage**
save to local database

**Step 5: Visualization**
show tweets in a map

3. A Python demo of Twitter data collection and processing

Twitter API

Twitter Warehouse

Developer/Researcher
3. A Python demo of Twitter data collection and processing

Twitter API

Who is it?

Twitter Warehouse

Developer/Researcher
3. A Python demo of Twitter data collection and processing

Twitter API

Twitter Warehouse

Developer/Researcher

I am Lei.

(1. Register a Twitter account)
3. A Python demo of Twitter data collection and processing

Twitter API

Twitter Warehouse

Developer/Researcher

What can I do for you?
3. A Python demo of Twitter data collection and processing

Twitter API

Twitter Warehouse

Developer/Researcher

I need some Twitter data.
(2. Create a new application)
3. A Python demo of Twitter data collection and processing

Twitter API

Here is the key. You can come in and start collecting data now.

(3. Obtain keys and tokens)
3. A Python demo of Twitter data collection and processing

Twitter API

https://apps.twitter.com/

1. Register a Twitter account
2. Create a new application
3. Obtain keys and tokens
   - Consumer Key (API Key)
   - Consumer Secret (API Secret)
   - Access Token
   - Access Token Secret
4. Save keys and tokens to a file
3. A Python demo of Twitter data collection and processing

Step 1: Collection

**Input:** Keywords, keys and tokens

**Run python code:** Step1_Collection.py

**Output:** printing raw twitter data containing defined keywords
3. A Python demo of Twitter data collection and processing

Step 2: Parsing

Data collected from Twitter API

In users’ eyes

In researcher/developer’s eyes
3. A Python demo of Twitter data collection and processing

Step 2: Parsing

Expand User Profile
3. A Python demo of Twitter data collection and processing

Step 2: Parsing

Run Python code: Step2_Parsing.py

Output: Printing tweets and locations
3. A Python demo of Twitter data collection and processing

Step 3: Geocoding

Google Maps API:
http://maps.googleapis.com/maps/api/geocode/json?address=UNT  (Input the Address)

```
"formatted_address" : "1155 Union Cir, Denton, TX 76203, USA",
"geometry" : {
  "location" : {
    "lat" : 33.207488,
    "lng" : -97.1525862
  },
  "location_type" : "ROOFTOP",
  "viewport" : {
    "northeast" : {
      "lat" : 33.2088369802915,
      "lng" : -97.15123721970851
    },
    "southwest" : {
      "lat" : 33.2061390197085,
      "lng" : -97.15393518029151
    }
  }
},
"place_id" : "ChIJyY815HPKTYYRCmlH-kWfvQs",
"types" : [ "establishment", "point_of_interest", "university" ]
```
3. A Python demo of Twitter data collection and processing

**Step 3: Geocoding**

Run Python code: Step3_Geocoding.py, Geocoding_Example.py

Output: printing pairs of coordinates (Longitude, Latitude)
3. A Python demo of Twitter data collection and processing

Step 4: Storage

Run Python Code: Step4_Storage.py

Output: tweetDB.csv with 50 pairs of coordinates
3. A Python demo of Twitter data collection and processing

Step 5: Analysis & Visualization

**Input:** tweetDB.csv (Data file generated in step 1-4)

**Run Python Codes:** Step5_Visualization.py, Step5_Heatmap.py

**Files Needed:** tweet_visualization.html, tweet_visualization_heatmap.html

**Output:** Two web maps

![Distribution of collected Tweets](image1)

![Heatmap of collected Tweets](image2)
Conclusion

1. Introduction of location-based social media data mining
   • Big Data, geographic information, volunteered contents in real-time
   • Marketing, survey, emergency management, etc

2. How to conduct location-based social media data mining?
   • The 5-step framework
   • Collection, Parsing, Geocoding, and Analysis

3. A Python demo of Twitter data collection and processing
   • Twitter API
   • Data collection and processing
   • Web mapping
Useful Resources

1. **Social Media Data Mining Books**
   - Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More. (Author: Matthew A. Russell)
   - Social Media Mining: An Introduction. (Authors: Reza Zafarani, Mohammad Ali Abbasi, Huan Liu)

2. **Article**

3. **Python programming**
   - [https://pythonprogramming.net/](https://pythonprogramming.net/)
   - Code for this demo: [www.rsgis.envs.lsu.edu/lzou/gisworkshop.html](http://www.rsgis.envs.lsu.edu/lzou/gisworkshop.html)
   - Your best partner – Google
Thinking Spatially for a Sustainable Future!

Lei Zou (lzou4@lsu.edu)
Website: www.rsgis.envs.lsu.edu/lzou