

Introduction to Social Media Analytics for Researchers

October 19th, 2017

Kang Pyo Lee, Ph.D.

UI3 / ITS-RS



About the Speaker

- **Name:** Kang Pyo Lee
- **Motto:** “Learn from data!”
- **Education:** Ph.D. in Computer Science at Seoul National University
- **Previous Work:** data scientist at Samsung Big Data Center
- **Current Work:** staff data scientist at UI3 and ITS-RS
- **Research Interests:** data science, social media analytics, big data, etc.

Why Social Media Analytics?

Q:

Why are researchers interested in analyzing social media?

A:

They want to understand the real world by looking at the social media world

Why Social Media Analytics?

Q:

Can we say the social media truly represents the real world?

A:

Not always! It could be **biased towards a specific group of people.**

Why Social Media Analytics?

Q:

**How can we overcome
the biases of social media?**

A:

Identify the users of each source
and choose the right sources

Market Research

Identify and analyze

- **market need**
- **market size**
- **competition**

from Wikipedia

Market Research as an Application Example

Traditional Market Research

Vs.

Big Data Based Market Research

Want to know about the true entire market → sampling!

Surveys, polls, and interviews

Social Media Analysis

Ask directly!

Infer from conversations!

Small sample (at most 1K subjects)

Large sample (thousands, millions, billions)

Market Research as an Application Example

**Traditional
Market Research**

Vs.

**Big Data Based
Market Research**

So, which one is better?

Market Research as an Application Example

**Traditional
Market Research**

Vs.

**Big Data Based
Market Research**

Both approaches are complementary

Choose the best approach for your research!

If you need brilliant ideas → find a few smart users

If you need ideas from many people → look at social media

Social media analysis is never easy because...

1. It requires **domain knowledge**
2. It deals with **human language**

Types of Analytics Applied to Social Media Analytics

- **Statistics** for understanding numbers
- **Text analytics** for understanding text
- **Network analytics** for understanding user networks
- **Geospatial analytics** for understanding geographical or spatial characteristics

Twitter has been known as the best source for social media analytics

- Very popular (still?)
- **Data friendly**
- Informational
- Short text → tweets
- Information dissemination → retweets
- Open networks → follow



Twitter Analytics

- **Real-time tweets** → **Streaming API**
<https://dev.twitter.com/streaming/public>
- **Historical tweets** → **Search API (1 week)**
<https://dev.twitter.com/rest/reference/get/search/tweets>
- **User networks** → **Followers API**
<https://dev.twitter.com/rest/reference/get/followers/ids>
- **Geolocation** → **Geo API**
<https://dev.twitter.com/rest/reference/get/geo/search>

Non-Twitter Analytics

- **Other social media services**
: Facebook, Instagram, Snapchat, ...
- **Blog posts**
- **E-commerce customer reviews**
- **Community posts**
- **News articles**
- **Search trends**

Data Collection

- APIs are always preferred over web scraping
 - APIs provide easier access to their data in a more structured manner
- Being able to collect data does NOT always mean that you're allowed to use the data at your disposal
 - The responsibility of the data falls on the user when collected
 - You must adhere to the Terms of Service of the service
- You must not violate the politeness policy that states how to avoid overloading web sites

Twitter Analysis Example



Analysis_Addiction_KPL.html



Analysis_Addiction_no_code_KPL.html