Social Media Computing

Lecture 1: Introduction

Lecturer: Aleksandr Farseev
E-mail: farseev@u.nus.edu
Slides: http://farseev.com/ainlfruct.html
Outline

• Age of Social Media
• Analysis of Social Media
• Challenges in Social Media
The New Information Age

- The Internet has revolutionized the way information is created, disseminated and consumed
  - Mixture of info available has changed from purely text, to include mm data, and live media
  - Emergence of huge amount of end-user generated data, especially in social networks
  - Greater connectivity leads to huge amount of live info
  - Internet has also rapidly gone mobile, permitting access from anywhere
How Big is Internet?

• The size is about 1 Trillion pages, but ~ 4.83 billion pages are indexed (http://worldwidewebssize.com, Jan 2014)

• Studies claimed that the deep web (the dynamic pages) is ~500 times larger than the indexed
  ▪ This does not include the huge amount of (real-time) forum and social postings

• Like deep Web, the amount of info available in live and social Web is huge
Traditional Media

Communication Media: One-One

Broadcast Media: One-Many
Social Media

Many-to-Many
Rich UGCs in Web 2.0

As of August 2014
Characteristics of Social Media

- Everyone can be a media outlet
- Rich User Interactions
- User-Generated Contents
- User Enriched Contents
- Collaborative environment
- Collective Wisdom
Characteristics of Social Media

Broadcast Media
Filter->Publish

Social Media
Publish->Filter

• How would these characteristics affect the ways we look into social media
Outline

• Age of Social Media
• Analysis of Social Media
• Challenges in Social Media
• Course Arrangements
Social Media Platforms

• Social Network platforms
  – Three major platforms: Private, Professional, Public

• Media Sharing platforms
  – YouTube, Instagram, Flickr, . . ., Vine

• Social Messaging platforms
  – Whatsapp, LINE, Wechat, SnapChat, . . .

• Social Curation platforms: Pinterest

⭐ The recent apps are all image/video based..
Users at Center of Social Media Environment

Public

Personal -- Friends
Functions of Social Media Platforms

• From users’ perspective
  – **Communications**: sharing, interacting, keeping up-to-date.. with friends
  – **Expression**: air/project views
  – **Self-preservation**: Wellness, exercise, self-improvements
  – **Local communities**

• From platform providers’ perspective
  – **Attract** and retain users; enrich contents; monetization
  – **Offer innovative/fun services**; improve user engagement
  – **Understand users**: co-viewing and co-creation of contents with users
NExT Social Observatory at NUS

- Gathering of multi-source data for integrated analysis

Types of UGC’s Gathered

**Type 1:** Contents: Tweets; Comments, cQA
- User
- Comments/cQA
- Twitter
- Social News

**Type 2:** Images/Videos
- Images/Videos

**Type 3:** Location/Check-in Apps
- Location Apps

**Type 4:** Structured Data
- Str. Data

Structured Contents

People, Domain, Social, Loc & Mobile

Users

Gathering of multi-source data for integrated analysis
Platform as a Service: Indian Elections

Sub-Events Around an Entity (Modi)

http://live.nextcenter.org
Platform as a Service: Indian Elections
Details of a Sub-Event (Modi)

http://live.nextcenter.org
Platform as a Service: Indian Elections
Relations between Entities (Modi)

http://live.nextcenter.org
Platform as a Service: Indian Elections
From Data to Event to Reports

Data Gathered from Twitter
INDIA ELECTIONS 2014
WEEKLY TRENDS
(Feb 24 - Mar 02)

Who will be the next PM of India?

ARVIND KEJRIWAL
AAP (Aam Aadmi Party)

NARENDRA MODI
BJP (Bharatiya Janata Party)

RAHUL GANDHI
INC (Indian National Congress)

Q1: How many mentions for the candidates?

Q2: What are the results of sentiment analysis on the candidates?
Which Social Platforms are the most influential?
ACTIVE USERS BY SOCIAL PLATFORM

MOST RECENTLY PUBLISHED MONTHLY ACTIVE USER ACCOUNTS BY PLATFORM, IN MILLIONS

<table>
<thead>
<tr>
<th>Platform</th>
<th>Social Network</th>
<th>Messenger / Chat App / VoIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>1,415</td>
<td>829</td>
</tr>
<tr>
<td>QQ</td>
<td>700</td>
<td>829</td>
</tr>
<tr>
<td>WhatsApp</td>
<td>500</td>
<td>629</td>
</tr>
<tr>
<td>Qzone</td>
<td>468</td>
<td>629</td>
</tr>
<tr>
<td>Facebook Messenger</td>
<td></td>
<td>500</td>
</tr>
<tr>
<td>WeChat</td>
<td>300</td>
<td>468</td>
</tr>
<tr>
<td>Skype</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td>Instagram</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td>Baidu Tieba</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td>Google+</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td>Twitter</td>
<td>300</td>
<td>288</td>
</tr>
<tr>
<td>Viber</td>
<td></td>
<td>236</td>
</tr>
<tr>
<td>Tumblr</td>
<td>230</td>
<td>236</td>
</tr>
<tr>
<td>Snapchat</td>
<td></td>
<td>200</td>
</tr>
<tr>
<td>Line</td>
<td></td>
<td>181</td>
</tr>
<tr>
<td>Sina Weibo</td>
<td></td>
<td>167</td>
</tr>
<tr>
<td>YY</td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>VKontakte</td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>LinkedIn</td>
<td></td>
<td>93</td>
</tr>
<tr>
<td>BBM</td>
<td></td>
<td>91</td>
</tr>
</tbody>
</table>
FaceBook

- Facebook is by far the largest social network site in the world, overtaking Google in 2007 in terms of user traffic.
- It is also the largest photo sharing site, with more than 250 million uploads each month (2012).
Twitter

- Twitter, a Microblogging site, is the second most popular with over 650 million users
- First popular English-speaking microblog
- Twitter messages are mostly public, whereas FB messages are mostly private, they serve different purposes
  - Twitter serving public forum
  - FB targets private groups
Characteristics of Twitter

- Its messages are restricted to **140 characters in length**:
  - Encourages users to post **short and frequent messages**
  - However, **messages are informal** and contains abbreviations
    Eg: u must be talkin bout the paper but I was thinkin movies
  - Hard to analyze by formal NLP tools

- It supports **social functions**, like follow, re-tweet & reply:
  - “Follow” permits someone to follow anyone; it is very popular and copied by most social network services
  - Re-tweets are widely used too

- It is conversational in nature:
  - Messages tend not to repeat terms used earlier, and
  - Vocabulary used change dynamically
User’s Tweeting Habits Over Time

- A concrete example

@Crushedsj 2013-10-17 13:15:24
don't sure yet, i'm trynna find a good medical school so probably nus

detail

@Crushedsj 2013-10-17 18:21:41
@AsiaPacNews: NUS professors to ride from Singapore to Sweden for breast cancer research
http://t.co/cNGTwedSsq #indonesia

detail

@Crushedsj 2013-10-17 20:41:48
My future school !! I swear I will take all the effort to get in here !!!

detail
Location-Based Social Networks

- Encourages users to "check-in" at various local venues, and enter comments and/or upload photos
- Allows many local places to be tagged
- Search support: find list of nearby places; whereabouts of friends, etc.

- Many other similar LBSNs, like Gowalla, Yelp etc..
Typical Activities on 4Square
Distributions of 4Square Venues

33 million users, 3.5 billion check-ins, as at April 2013

40+ million users, 4.5 billion check-ins, as at September 2013
Gardens by the Bay brings to life NParks’ vision of creating a City in a Garden. The Gardens captures the essence of Singapore as the premier tropical Garden City with the perfect environment in which to live and work - making Singapore a leading global city of the 21st century.
Mobile Photo Sharing Sites - Instagram

- Fun app: give users capability to transform picture to more professional looking ones
- Bought by Facebook for US$1 Billion
Visual-based Social Media Platforms

- Many emerging popular applications are visual based
  - **Pinterest**: founded in 2010; popular with women and is very targeted (~70M users)
  - **Vine**: founded in 2012; sharing of 6-sec loop-videos; popular with young people (~40M users)
  - **Snapchat**: founded in 2011; Sharing 10-sec video moments that erases itself after certain period; 70% of users <24 yrs old (~100M active users); user base will be bigger than Twitter soon
Social Networks between People

• One key aspect of social network systems is the network relationships between users

• Social Network: A social structure made of vertices (individual users) that are related to each other implicitly (similar behavior) or explicitly (friendship)

• Graphical Representation:
  – Vertices => members
  – Edges => relationships
    • Adjacency Matrix/List
Outline

• Age of Social Media
• Analysis of Social Media
• Challenges in Social Media
A Recap: Key Characteristics of Current Generation of Social Media Systems

- **Live, Big-Data and 3M** (Multi-source, Multimedia & Multilingual)
- **Support users’ need for communication**, sharing and interaction
- **Support co-viewing and co-creation of contents** (by users and systems)
- **Develop social analytics** aim to understand contents, events and users - targeting at recommendation
What Has Changed in Last 5 Years?

1. Image/Video handling
   - Top 3 recent social media platforms are all image/video centric
   - Live video from GoPro

2. Live Location Analytics
   - Sensor devices are everywhere – capable of multi-form of sensing
   - Multi-sources info: Location traces, POI and Audio
   - Towards better location estimation and mobility analytics.
What Has Changed in Last 5 Years

3. Online and Live are Central
   – *Live*: comes the ability for continuous sharing, interactions and feedback.
   – Users want to get instant feedback – from friends and systems

4. Quality and Structures of Data
   – Deteriorating quality of data, with about 70% of UGCs belongs to noise/spam/rumors category
   – Key part of making data usable is to structure them: at both knowledge and data level

5. Co-Creation & Co-Invention
   – With live instant feedback, comes the possibility to co-create and co-invent
   – Not just the contents, but systems and design
Summary

• We live in new era of Social Media. It is really Big and full of heterogeneous relations of type “Many – Many”

• The data is not just Big, but multi-source and of different modality:
  – Text from Microblogs and other Social Forums
  – Location From Location-Based Social Networks
  – Images from Image Sharing Services
  – Video from Video Sharing Services

• We live in new era of Social Media. It is really Big and full of heterogeneous relations of type “Many – Many”

• Data become more noisy, New types of data emerge:
  – Live data
  – Sensor data
  – Etc.
# Course Schedule

<table>
<thead>
<tr>
<th>Wk</th>
<th>Date</th>
<th>Lecture/Tutorial Topics</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>9 Nov</td>
<td>L1: <strong>Introduction</strong> to Social Networks &amp; Challenges</td>
<td>Details of Assgn 1</td>
</tr>
<tr>
<td>2.</td>
<td>9 Nov</td>
<td>L2: <strong>Text</strong> Processing</td>
<td>T1: Text representation</td>
</tr>
<tr>
<td>3.</td>
<td>9 Nov</td>
<td>L3: <strong>Location</strong> and <strong>Image</strong> Data Processing, Clarification on Assign. 1</td>
<td>T2: Location and Image data representation</td>
</tr>
<tr>
<td>4.</td>
<td>10 Nov</td>
<td>L4: Introduction to <strong>Retrieval</strong> and <strong>Classification</strong></td>
<td>T3: Introduction to Information Retrieval and Classification</td>
</tr>
<tr>
<td>5.</td>
<td>10 Nov</td>
<td>L5: <strong>Source Fusion</strong> and Evaluation</td>
<td>T4: Data Source Fusion</td>
</tr>
<tr>
<td>6.</td>
<td>10 Nov</td>
<td>L6: Recent Study</td>
<td>T5: Additional Topics in Social Media</td>
</tr>
<tr>
<td>7.</td>
<td>12 Nov</td>
<td>L7: Wrap up</td>
<td>Summary of the above</td>
</tr>
<tr>
<td>8.</td>
<td>12 Nov</td>
<td><strong>Group presentations of the Assignment 1</strong></td>
<td>Assign 1 Due</td>
</tr>
</tbody>
</table>
Assessment:

• 50%: Lecture Participation
• 50%: 1 Programming Assignment
Assignment -1

- In addition to theory and knowledge, one key aspect of this course is hands-on exercises

- One assignment (groups of 3):
  
  - **Prediction of User Demography**
    
    Given info from *multiple social media platforms* from *three geographical regions*, determine the demography of users such as:
    
    - Age
    - Gender
    - Education Level
    - Occupation Industry
    - Relationship Status.

  - 10 min. group presentations of projects after the last lecture on **Thursday Nov. 12, 2015**
Assignment -2

• Presentation should be 15 min long include:
  – Team members and their roles introduction
  – Evaluation Results:
    • In terms of MACRO Precision, MACRO Recall, MACRO F-Measure
    • Based on those users, who have mentioned their real age (In ground truth files, the column “Real Age” is not empty) and for every city.
Assignment -3

• DATASET:
  http://lms.comp.nus.edu.sg/research/NUS-MULTISOURCE.htm

• DESCRIPTION OF THE DATA IS IN PAPER*.

• Please, ask any questions during the conference and after: farseev@u.nus.edu

Assignment -4

• All slides and will be here:
  – http://farseev.com/ainlfruct.html

• Recommended software to use:
  – KNIME (No programming required)
    https://www.knime.org/
  – Python and it’s Machine Learning Support
  – Any other language you like. Just make it work ;)


http://next.comp.nus.edu.sg/opportunities

- RESEARCH INTERN
- ARCHITECT / JAVA DEVELOPER
Next Lesson

• Text Processing