E6895 Advanced Big Data Analytics Lecture 2:

*Foundations for Human-Like AI*

Ching-Yung Lin, Ph.D.
Adjunct Professor, Dept. of Electrical Engineering and Computer Science
IBM Chief Scientist, Graph Computing

January 27, 2017
Computational Discovery of Social Cognitive Essence

Actionable Applications
- Anomaly Detection
- Live Monitoring
- Flow Manipulation
- Predictive Visualization
- Auto-Counter Messaging
- Intranet-Social-Media Action

Inferred Cognitive Traits
- (Human Essential)
  - Personality
  - Needs
  - Value
  - Trustworthiness
- (Human Dynamic)
  - Contextual Behavior
  - Emotional State
- (Information Dynamic)
  - Info Reasoning & Morphing
  - Visual Sentiment

Inferred Social Network Traits
- Roles
- Dynamic Analysis
- Topological Analysis
- Location Analysis

Analytics & Predictive Models
**Goal 1:** Detect, classify, measure and track the
(a) formation, development, and spread of ideas & concepts (memes)
(b) purposeful or deceptive messaging and misinformation

**Goal 2:** Recognize persuasion campaign structures and influence operations across social media sites and communities

**Goal 3:** Identify participants and intent, and measure effects of persuasion campaigns

**Goal 4:** Counter messaging of detected adversary influence operations

---

Example of our prior work on cognitive understanding through social media:

53+ papers published, accepted, & submitted
12+ patents filed
ACM CIKM 2012 Best Paper Award
IEEE BigData 2013 Best Paper Award
PNAS Cover Article Jan 2013
Science (1)
Nature (2)
Social Media Data

- Memes Persuasiveness in Multiple Networks
  - Social Content Analysis
    - Topic Modeling
    - Information Morph
  - Content models

- Personality Analysis
  - Behavioral Modeling
    - Emotion Analysis
    - Location Analysis
  - Behavioral models

- Role Discovery
  - Network Controllability & Observability
  - Social Capital Modeling
  - Social network models

- Visual Manipulation Modeling & Detection
  - Visual memes evolution models

Social Media Applications

- Social Malware Detection
- Affecting Memes Propagation
- Counter Messaging Planning

Real-time Large-scale Social Media Mining Algorithms

Real-time Large-scale Social Media Mining System

UI & Visualization

Social Mining Architecture
Our example prior work on cognitive understanding through social media

**Thrust 1. Modeling Information Dissemination in Context:**
- Task 1.1. Computational Modeling of User Dynamic Behavior
- Task 1.2. Computational Models of Trust and Social Capital
- Task 1.3. Information Morphing Modeling
- Task 1.4. Persuasiveness of Memes
- Task 1.5. The Observability of Social Systems
- Task 1.6. Culture-Dependent Social Media Modeling
- Task 1.7. Dynamics of Influence in Social Networks
- Task 1.8. Understanding the Optimal Immunization Policy
- Task 1.9. Modeling and Identification of Campaign Target Audience
- Task 1.10. Modeling and Predicting Competing Memes

**Thrust 2. Detecting and Tracking Information Dissemination in Context:**
- Task 2.1. Real-Time and Large-Scale Social Media Mining
- Task 2.2. Role and Function Discovery
- Task 2.3. Detecting Malicious Users and Malware Propagation
- Task 2.4. Emergent Topic Detection and Tracking
- Task 2.5. Detecting Evolution History and Authenticity of Multimedia Memes
- Task 2.6. Synchronistic Social Media Information and Social Proof Opinion Mining
- Task 2.7. Community Detection and Tracking
- Task 2.8. Interplay Across Multiple-Networks

**Thrust 3. Affecting Information Dissemination in Context:**
- Task 3.1. Crowd-sourcing Evidence Gathering to Formulate Counter-messaging Objectives
- Task 3.2. Delivery and Evaluation of a Counter-messaging Campaign
- Task 3.3. Optimal Target People Selection
- Task 3.4. Automated Generation of Counter Messaging
- Task 3.5. User Interfaces for Semi-Automatic Counter Messaging
- Task 3.6. Controlling the Dynamics of Influence in Social Networks
- Task 3.7. Influencing the Outcome of Competing Memes and Counter Messaging
Inferring Cognitive Traits: Human Essentials

- Personality
- Needs
- Value
- Trustworthness / Trustingness
- Influence
Deriving Personality

Big5 Personality (OCEAN)

- Extraversion: outgoing/energetic vs. solitary/reserved
- Openness: inventive/curious vs. consistent/cautious
- Conscientiousness: efficient/organized vs. easy-going/careless
- Agreeableness: friendly/compassionate vs. cold/unkind
- Neuroticism: sensitive/nervous vs. secure/confident
Deriving Personality

- Mapping text to psycholinguistic category (LIWC) to BIG 5 Personality [Yakoni '10]: 694 bloggers; 66 LIWC categories; ~2,500 words

- We extended the # of words to about 30,000 by combining with WordNet

<table>
<thead>
<tr>
<th>LIWC Category</th>
<th>N</th>
<th>E</th>
<th>O</th>
<th>A</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total pronouns</td>
<td>0.06</td>
<td>0.06</td>
<td>0.21***</td>
<td>0.11**</td>
<td>-0.02</td>
</tr>
<tr>
<td>First person sing.</td>
<td>0.12**</td>
<td>0.01</td>
<td>-0.16***</td>
<td>0.05</td>
<td>0</td>
</tr>
<tr>
<td>First person plural</td>
<td>-0.07</td>
<td>0.11**</td>
<td>-0.1*</td>
<td>0.18***</td>
<td>0.03</td>
</tr>
<tr>
<td>First person</td>
<td>0.1*</td>
<td>0.03</td>
<td>-0.19***</td>
<td>0.08*</td>
<td>0.02</td>
</tr>
<tr>
<td>Second person</td>
<td>-0.15***</td>
<td>0.16***</td>
<td>0.12**</td>
<td>0.08</td>
<td>0</td>
</tr>
<tr>
<td>Third person</td>
<td>0.02</td>
<td>0.04</td>
<td>-0.06</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>Negations</td>
<td>0.11**</td>
<td>-0.05</td>
<td>0.13**</td>
<td>-0.03</td>
<td>-0.17***</td>
</tr>
<tr>
<td>Assent</td>
<td>0.05</td>
<td>0.07</td>
<td>-0.11**</td>
<td>0.02</td>
<td>-0.09*</td>
</tr>
<tr>
<td>Articles</td>
<td>-0.11**</td>
<td>-0.04</td>
<td>0.2***</td>
<td>0.03</td>
<td>0.09*</td>
</tr>
<tr>
<td>Prepositions</td>
<td>-0.04</td>
<td>-0.04</td>
<td>0.17***</td>
<td>0.07</td>
<td>0.06</td>
</tr>
</tbody>
</table>

- Trait: Neuroticism
  - No. of words sig. at p < 0.01: 24
  - Top 20 words: awful (0.26), though (0.24), lazy (0.24), worse (0.21), depressing (0.21), irony (0.21), road (-0.2), terrible (0.2), Southern (-0.2), stressful (0.19), horrible (0.19), sort (0.19), visited (-0.19), annoying (0.19), ashamed (0.19), ground (-0.19), ban (0.18), oldest (-0.18), invited (-0.18), completed (-0.18)

- Trait: Extraversion
  - No. of words sig. at p < 0.01: 20
  - Top 20 words: bar (0.23), other (-0.22), drinks (0.21), restaurant (0.21), dancing (0.2), restaurants (0.2), cats (-0.2), grandfather (0.2), Miami (0.2), countless (0.2), drinking (0.19), shots (0.19), computer (-0.19), girls (0.19), gloriously (0.19), minor (-0.19), pool (0.18), crowd (0.18), sang (0.18), grilled (0.18)

- Trait: Openness
  - No. of words sig. at p < 0.01: 393
  - Top 20 words: folk (0.32), humans (0.31), of (0.29), poet (0.29), art (0.29), by (0.28), universe (0.28), poetry (0.28), narrative (0.28), culture (0.28), giveaway (0.28), century (0.28), sexual (0.27), films (0.27), novel (0.27), decades (0.27), ink (0.27), passage (0.27), literature (0.27), blues (0.26)

- Trait: Agreeableness
  - No. of words sig. at p < 0.01: 110
  - Top 20 words: wonderful (0.28), together (0.26), visiting (0.26), morning (0.26), spring (0.25), porn (-0.25), walked (0.23), beautiful (0.23), staying (0.23), felt (0.23), cost (-0.23), share (0.23), gray (0.22), joy (0.22), afternoon (0.22), day (0.22), moments (0.22), hug (0.22), glad (0.22), fuck (-0.22)

- Trait: Conscientiousness
  - No. of words sig. at p < 0.01: 13
  - Top 20 words: completed (0.25), adventure (0.22), stupid (-0.22), boring (-0.22), adventures (0.2), desperate (-0.2), enjoying (0.2), saying (-0.2), Hawaii (0.19), utter (-0.19), it's (-0.19), extreme (-0.19), deck (0.18)
Deriving Needs

What do we model

- 12-dimension needs

[Ford, 2005]
Deriving Value

- Why model value
  - Values motivate people and guide their actions
  - Values transcend specific actions and situations

[Schwartz 2006]
Example of Personality/Needs/Value/Behavior
• **Trustingness:** How likely an actor is to trust another actor in the network.
  - A highly trusting actor trusts a lot of non-trustworthy actors
  - Higher score = Higher trustingness

• **Trustworthyness:** How likely an actor is to be trusted by others in the network.
  - A highly trustworthy actor is trusted by lots of non-trusting actors
  - Higher score = Higher trustworthyness

• 2 measures are negatively co-related to one another and are dependent on one another
  - Based on Hubs and Authority model from HITS algorithm
  - Twitter is a social media platform where people can verify/know identities of real persons leading to high trustworthy scores
Inferring Cognitive Traits: Human Dynamics

– Contextual Behavior
– Emotional State
Computational Modeling of Contextual Behavior

- **Objective:** Modeling user dynamic behavior for prediction/detection tasks

- **Task Goals:**
  - Modeling synchronous behavior at multiple granularities
    - Predict values (e.g., performance, credibility) of entities in heterogeneous networks
    - Published at SDM 2013
  - Modeling user dynamic information spreading behavior (ongoing, in demo)

- **The Work:**
  - Exploit the structure of people’s dynamic behaviors to facilitate prediction tasks, while few existing approaches consider that
Collective Intelligence and System Change Prediction

Data:
- All instant messages and trades by employees of a large hedge fund.
- 24 are traders, 95 are analysts, 63 are portfolio managers, 8646 outside contact
- 47K trades
- 22 million IMs (2008 – 2012)

Findings: We identify two behavioral patterns that signal system changes:
- Reaction to IMs containing relevant information
- In-group vs. out-group communication

Using these two features, we can make predictions of system-wide changes with better accuracy than the group’s predictions.
13:11:33, I was thinking all this AAPL anti-trust might be actionable
13:11:42', not great for AAPL
13:11:47, When GOOG had that big issue in Europe stock underperformed right?
13:11:52, true
13:14:01, Also not sure if you caught, but GSCO is going to allow employees to bring own phone device for corporate email
13:14:24, Maybe GSCO allowing that could be positive for AAPL, as security focused firm saying iPhone works
13:14:35, But bad for RIMM
13:14:42, Maybe all this is priced in
13:16:50, Did you see speculation that Bing is actually quietly going to be default search on iPhone 4?
13:17:18, heard a lot of talk of that
13:17:23, but didn't see that specifically like that
13:17:44, Okay let me figure out where I saw that and get back to you
13:17:45, One sec
13:20:08, [Link](http://thenextweb.com/apple/2010/06/07/wait-bing-is-default-search-on-iphone-4)
**Motivation**
Emotional states can effect how information is processed. Good information can be undermined or strengthened by emotional states.

**Approach**
- Measure emotional activation in tweets using the ANEW dictionary
- Control for the number of words in the text

**Preliminary experiments**
- Traders at a hedge fund are more likely to make decision errors when they are very emotionally activated or very emotionally deactivated.
- Users who retweeted the 20 detected most anomalous sequences tend to post tweets with higher level of emotion than a baseline of 20 million tweets from June 2009.
The Power of Social (Visual) Multimedia
- A picture is worth one thousand words

Tweets of the Year -- 2012

@BarackObama: Four more years.    @Brynn4NY: Rollercoaster at sea.
Question: How to Build Visual Sentiment Ontology?

-- Web + big data + computer vision + psychology

Psychology emotion wheel
(24 emotions, by Robert Plutchik)

Plenty on the Web:
“For content to go viral, it needs to be emotional,” Dan Jones
Research: Which 1000 sentimental concepts?
-- data mining to discover visual sentiments in social media

Psychology emotion wheel (24 emotions)

Build Sentiment Ontology

Select Adj-Noun Pairs

Analyze tags with strong sentiments

MISTY WOODS
Beautiful Sky

Beautiful Flower
**Example of Robot -- RoboCup Introduction**

RoboCup is an international robotics competition founded in 1997.

**RoboCupSoccer Competitions**
- Simulation League
- Small Size League
- Middle Size League
- Standard Platform League
- Humanoid League

24 qualified teams participate in the Humanoid League

RoboCup2010, Singapore.

Speaker: Dr. Guangnan Ye
How was it played? (video)

Motion Ability
Competition Rules

The match lasts two equal periods of 10 minutes. A match is played by two teams, each consisting of not more than three players. One of the players must be designated as goalkeeper.

The Field of Play
The Framework of the Robot

- Camera
- Loudspeaker
- Wireless Network
- Microphone
- PC104
  - USB2.0
  - RX232
  - RX485
- Mega128
- Gyro
  - I2C
- Accelerator
- RX-28
- RX-28
- RX-28
- RX-28
The Architecture of the Robot

Fig. Design Picture

Fig. Physical Sample
Gait Planning

Kinematics Formulation

\[
0^T_6 = 0^T_1 1^T_2 2^T_3 3^T_4 4^T_5 5^T_6 = \begin{bmatrix}
a_{x1} & b_{x1} & c_{x1} & p_{x1} \\
a_{y1} & b_{y1} & c_{y1} & p_{y1} \\
a_{z1} & b_{z1} & c_{z1} & p_{z1} \\
0 & 0 & 0 & 1
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 & p_{x1} \\
0 & 1 & 0 & p_{y1} \\
0 & 0 & 1 & p_{z1} \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

\[
6^T_12 = 6^T_7 7^T_8 8^T_9 9^T_10 10^T_11 11^T_12 = \begin{bmatrix}
a_{x2} & b_{x2} & c_{x2} & p_{x2} \\
a_{y2} & b_{y2} & c_{y2} & p_{y2} \\
a_{z2} & b_{z2} & c_{z2} & p_{z2} \\
0 & 0 & 0 & 1
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 & p_{x2} \\
0 & 1 & 0 & p_{y2} \\
0 & 0 & 1 & p_{z2} \\
0 & 0 & 0 & 1
\end{bmatrix}
\]
Gait Planning

Fig. Software Interface for Gait Planning

Fig. Example of the Robot Walking
Goal and Ball Detection

a) Original Picture

b) Color Segmentation

c) Blob Formation

d) Post-processing
Localization

a) Original Image
b) Projection
c) Color Classification
d) Edge Detection

e) Play Field Detection
f) Line and Circle Detection
g) Calibration
h) Localization
Particle Filter Based Localization

- a) Robot Direction
- b) Particles Formation
- c) Weight Formation
- d) Motion Update
- e) Observation Update
- f) Re-sampling
Existing Software Tools example – Choregraphe (for NAO)

SDKs also contain a range of functions to help you learn about programming (Python, C++, Java, JavaScript) and develop your own behaviours with [NAO](https://www.ald.softbankrobotics.com/en/cool-robots/nao)