

Understanding Country Sentiment Using On Line Social Media

Clayton Fink, Victor McCrary
Science & Technology Business Area
victor.mccrary@jhuapl.edu
240-228-1763

The logo for Applied Physics Laboratory (APL) consists of the letters 'APL' in a large, bold, red, sans-serif font.

The Johns Hopkins University
APPLIED PHYSICS LABORATORY

Traditional Methods For Understanding Regional Opinion

- Official contacts
- Personal relationships
- Intel
- Traditional media
- Traditional polling and surveys

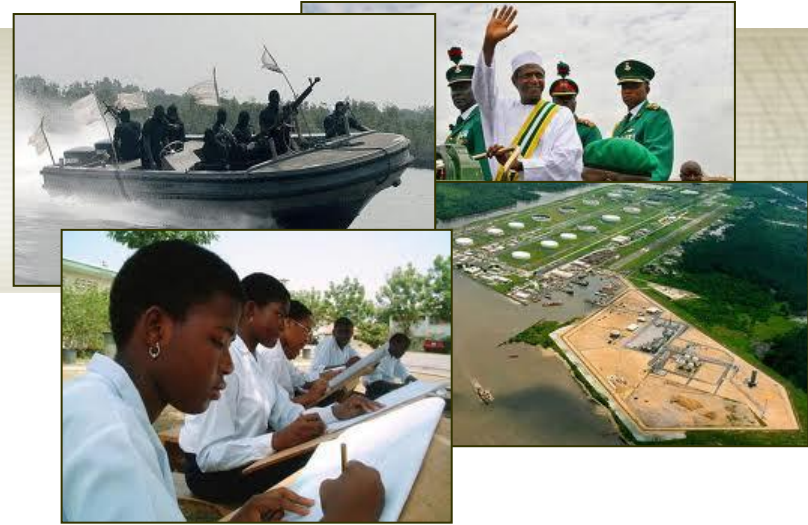


Are Social Media Useful For Gathering Opinion?

- Enormous growth in parts of the developing world
- Forum for expression of opinion
- Real time
- Captures connections between people



Focus on Nigeria



- Most populous country in Africa (158M)
- 2nd largest economy in Africa
- 4th largest source of oil for US
- Democracy, though suffers from endemic corruption
- Fledgling technology sector
- Active insurgency in Niger Delta
- Ethnic and religious tensions

Collecting Social Media Data From Nigeria

- Significant growth in social media use – currently has 3M Facebook users
- Collecting user generated text from multiple sources
- Most text is in English with some Nigerian Pidgin
- “Noisy data” with discussion of many topics...including politics



Social Media as a Source of Public Opinion

- **Example: Tweeting the 2011 Nigerian Presidential Campaign:**

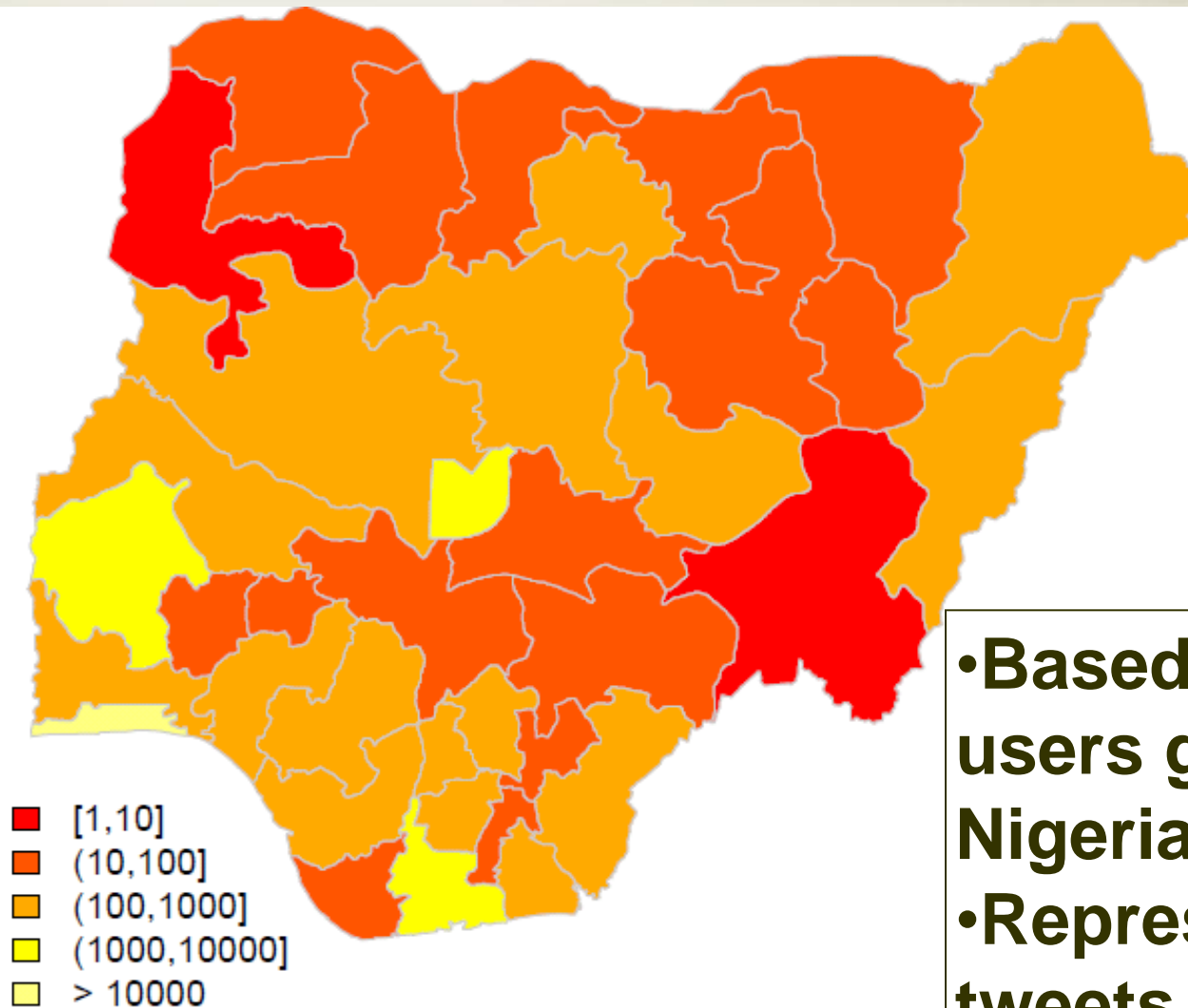
*If **Ribadu** has only one vote I can swear it's mine*

*@Mr_E I too am a believer in **Buhari** I totally agree with u.*

***Goodluck jonathan** will turn nigeria to a better place please let vote for him*

- **Using supervised machine learning to extract sentiment from text**
- **Also, how do we characterize the online population of opinion holders?**

Data Curation – Identifying Twitter Users By State



- Based on 51K users grounded to Nigeria
- Represents 23M tweets from Nigeria

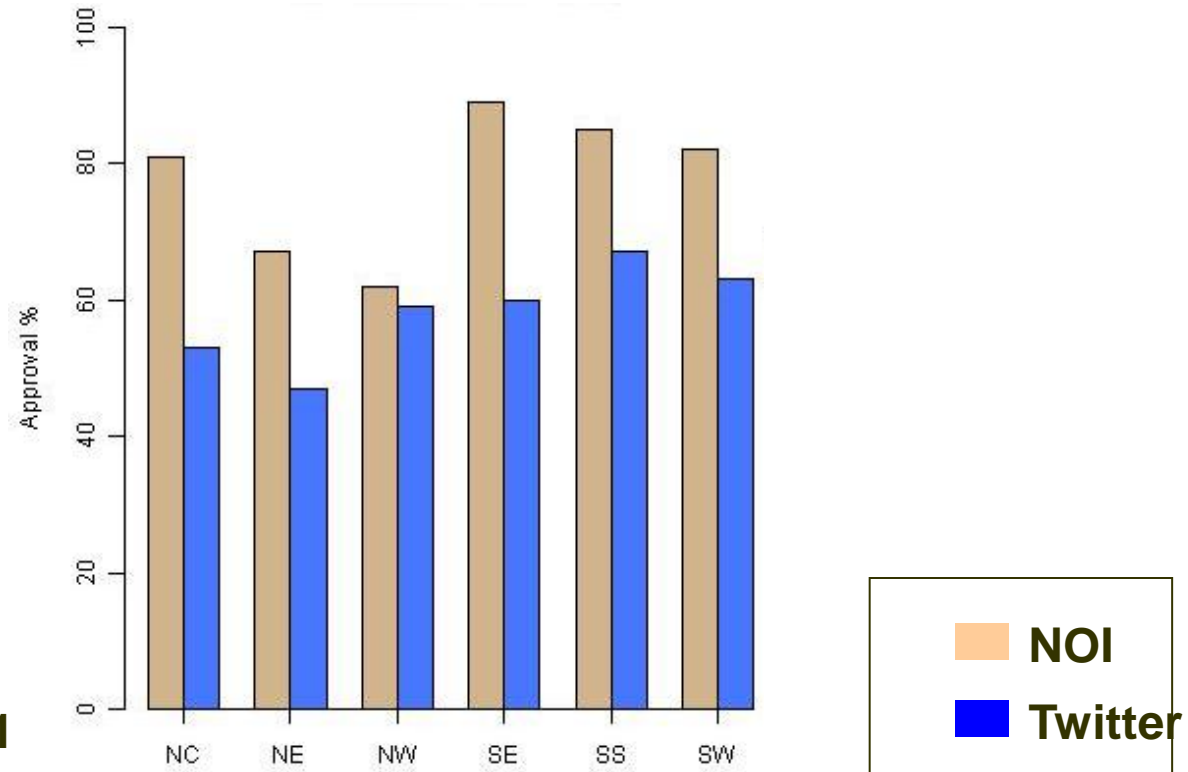
Data Curation – Automating Ethnicity *and* Gender Detection of Facebook Users

- **Used hierarchical Bayesian models to detect gender and ethnicity for Facebook users from Nigeria**
- **Used letter n-gram features of names and content-based features**
- **Accuracy: 80.1 for gender, 81.1 for ethnicity (Igbo and Yoruba only)**

Rao, D., Paul, M., Fink, C., Yarowsky, D., Oates, T., & Coppersmith, G (2011). Hierarchical Bayesian Models for Latent Attribute Detection in Social Media. ICWSM 2011

Baseline Sentiment Analysis of Tweets Using Lexical Clues

Approval of President Goodluck Jonathan by Region Compared to August 2010 NOI Poll



Correlation = 0.51