Mining and Analyzing Data from Social Media for Pragmatic Research

Bethany Kwan, PhD, MSPH; Jenna Reno, PhD

University of Colorado Anschutz Medical Campus

Learning Objectives

1. Identify audiences and potential uses of social media in pragmatic research
2. Identify approaches to mining data from social media and the web for research
3. Describe quantitative and qualitative analysis methods appropriate for social media data

Use of Social Media in Pragmatic Research

Implementation and conduct of research

- Stakeholder and community engagement or "citizen science"
- Dissemination and messaging channels
- Recruitment and outreach

Source of data for research

- "Secondary use"
- Communication research
- Network analysis
- Ethnographic research
- Public health surveillance
- Patient-generated health outcomes data

Mining Data from Social Media

- Manual approaches
- Connection via an Application Programming Interface (API)
  - Free on Twitter – search and download tweets (but limited to 1% of tweets)
- Third party vendors
  - Licensed with the platform for broader access (can be expensive)
  - Symplur: https://www.symplur.com/products/signals/
  - Social listening tools
- Named entity recognition and normalization (automated and manual)
  - Named entity recognition: identification of entities such as drugs, diseases, and medical events
  - Normalization: Mapping to predefined categories or standard medical ontologies
  - Dictionary lookup
- Text mining techniques (extracting features of free-text for further analysis)
  - N-gram, word embedding, sentence-dependency-based parse tree, Latent Dirichlet Allocation (LDA) topic modeling

Analysis of Data from Social Media

- Network analysis
- Qualitative content analysis
- Supervised and unsupervised machine learning
- Hypothesis testing
- Facebook message testing

Activity #1: Know Your Audience
- Pick one of the "uses of social media for pragmatic research"

- State a specific hypothetical or real example of how you might use social media in one or more ways in your research

- Who is your audience?
    - Consider adopters, influencers, potential saboteurs

- Where might you find this audience on social media?

- How do they use social media?

- Who are the influencers on social media?

- How might you partner with existing online communities?

Activity #2: Social media data mining and analysis plan
- Consider the audience, social media platform, and research topic you considered in Activity #1.

- What data types might be available from that social media platform?
    - Text data
    - Structural data
    - Metadata
    - Other?

- How might you mine the data?

- How might you analyze the data?

- Who do you need on your team?

References and Resources

1. Ru B, Yao L. A literature review of social media-based data mining for health outcomes research. Social Web and Health Research. 2019:1-4.
2. Taylor J, Pagliari C. Mining social media data: how are research sponsors and researchers addressing the ethical challenges? Research Ethics. 2018 Apr;14(2):1-39.

COPRH Con
Colorado Pragmatic
Research in Health
Conference

ACCORDS
ADULT AND CHILD CONSORTIUM FOR HEALTH OUTCOMES
RESEARCH AND DELIVERY SCIENCE
UNIVERSITY OF COLORADO | CHILDREN'S HOSPITAL COLORADO

Colorado Clinical and Translational
Sciences Institute (CCTSI)
UNIVERSITY OF COLORADO **DENVER | ANSCHUTZ MEDICAL CAMPUS**