Abstract

Purpose:
- Vaccines are critical technology to control disease spread, but vaccine refusal compromises efforts towards this end.
- Recent studies focus on the social rationales underlying vaccine refusal. However, effective techniques to address these rationales remain unclear.
- We use big data to analyze these sociotechnical interactions by studying the distribution of the public’s rationales regarding vaccine technology in the US.
- The journal Vaccine indicated with a special that social media influence patients’ behaviors related to vaccines.
- And, Twitter has proven effective in disease surveillance [2].
- Therefore, we use Twitter to collect and track social trends regarding vaccine refusal.

Data Source and Methods:
- To obtain social information about vaccine technology, we employ machine learning processes to filter Twitter data described in previous work [3]. This yields:
  a) If a tweet is relevant to vaccination or not;
  b) If a vaccine tweet is neutral or non-neutral sentiment;
  c) If a non-neutral vaccine tweet is of positive or negative sentiment.
- We also ensure tweets are geotagged using machine learning [4].

Analyses:
- Temporal, spatial, and social:
  - Use machine learning to detect topics automatically
  - Specific vaccination sub-conversations within various communities
  - Maps of the US detailing in which topics are discussed
  - Automatic identification of social media users’ demographics
  - e.g. age, gender, parent status, income, education [8]
  - By user: maps of the US detailing who discusses vaccination and where
- Combining this information will allow us to pinpoint how vaccine refusal rationales vary across and within communities.
- In addition to when and where they vary, change, or appear.

End goal:
- Such analyses will contribute towards understanding the interactions between the spread of a disease, the associated social dynamics, and how technology may be used to reduce overall disease prevalence.

Data Source: Social Media

- Data collected beginning from November 2014
- Measles outbreak in Disneyland prominent example of vaccination debate
- Cut off at March 2015 for preliminary analyses
- 2M+ vaccine-relevant tweets
- Many neutral, but more negative than positive

Methods and Analyses

Relevance and Sentiment:
- Support Vector Machines, Python + sklearn, modified parameters on precision-recall curves
- Favored precision over recall when necessary (e.g. neutral vs non-neutral sentiment)

Location augmentation:
- Carmen system infers locations using geocoding tools and automatic and manual alias resolution [3]
- Initial foray yielded 44% data with location info (limitations section for improvements)
- Where available: lat-long, city, county, state, country

Topics:
- Latent Dirichlet Allocation run over all vaccine tweets for 50 topics

Tables:

<table>
<thead>
<tr>
<th>Topic ID</th>
<th>Parameter</th>
<th>Category</th>
<th>Description / keywords of topics</th>
</tr>
</thead>
</table>
| 30       | 0.00521   | pro      | vaccine, kill, AMR, flu, good, debate, vaccination, anti, children, love, jen
| 15       | 0.00501   | pro      | herd immunity important anti, people, measles, autism, science, vaccines, pro, good, movement, step, stupid, outbreak, bad, viewpoint, herd, important, lie, immune |
| 39       | 0.00325   | pro      | vaccine, science, truth, ecf Holocaust, autism, media, facts, terrorism, debate, doctors, health, pro, trump, science, research, reality, vaccinoworker |
| 8        | 0.00292   | pro      | vaccinated, measles, mcvv, crip, preventable, polio, risk, recently unvaccinated |
| 24       | 0.00263   | pro      | vaccine, ebov, trial, monkey, effective, flu, high, promise, success, show, world, experimental, hope, development |
| 20       | 0.00396   | anti     | ecf Holocaust, vaccine, autism, old, fraud, autismwarriors, agefautism, blackmewscarry |
| 46       | 0.00231   | anti     | vaccines, kill, california, mandatory, exemptions, children, parents, medical, school, religious, forced, mandate, personal, senate, freedom, require, laws, tyranny, eliminate |
| 9        | 0.00215   | anti     | vaccinate, kids, parents, school, doctor, refusal, unvaccinated, crippled, slope, disneyland, politics, Neglectees, love, diseases, child, catch, admit, eliminated, choose, absolutely, thewastegofnd-artificial |
| 34       | 0.00209   | anti     | vaccinate, measles, autism, children, endgame, marxists, cdc Holocaust, mezzo, risk, infant, rmm, death, bad, kill, lie, vaccinatesvrkcls |

Analyses – Cont’d

Spatial:
- ArcGIS 10.2, Getis-Ord Gi* (statistical hotspot analysis) [6]
- Topic 46 – anti: gov’t mandate
- Topic 34 – anti: autism and adverse reactions in kids

Limitations / Improvements:
- Group and analyze by available granularity of location
- Inspect and deal with mistaken or mislabeled data
- Improve topics on subsets of tweets, ‘de-trend’ the general debate
- Supervised topic models
- Incorporate demographics

References