Supplementing Obesity-Related Surveillance with Persistent Health Assessment Tools

Meredith Keybl*, John Henderson, Guido Zarrella, John Gibson and Maeve Kluchnik

MITRE Corporation, Bedford, MA, USA

Objective

We developed Persistent Health Assessment Tools, PHAT, to equip public health policy makers with more precise tools and timely information for measuring the success of obesity prevention programs. PHAT monitors social media to supplement traditional surveillance by making real-time estimates based on observations of obesity-relevant behaviors.

Introduction

In response to the rise in obesity rates and obesity-related healthcare costs over the past several decades, numerous organizations have implemented obesity prevention programs. The current method for evaluating the success of these programs relies largely on annual surveys such as the Centers for Disease Control and Prevention's Behavioral Risk Factor Surveillance System (BRFSS) which provides state-by-state obesity rates. As a result, public health policy makers lack the fine-grained evaluation data needed to make timely decisions about the success of their obesity prevention programs and to allocate resources more efficiently.

Methods

We developed a practical interface that enables policy makers to leverage social media. Specifically, we developed models for predicting obesity rates from sets of tweets and developed a dashboard to provide interactive navigation and time slicing. We isolate tweets of interest and import them into the dashboard where policy analysts can query and browse to observe instances where a program's outreach was successful and identify ways it can be improved. These tools enable policy makers to study relevant demographics, adjust content and messaging faster, and closely measure their program's impact.

Model: Traditionally, survey data is imputed onto different populations using multilevel regression with post-stratification[1]. Our approach built models predicting a survey result from non-surveillance indicator variables - namely text in social media posts, location, and time - by fitting margin-matching models. Those models were then used to predict the missing results on other demographic slices from that same data. We used 2011 BRFSS data[2] state-aligned with a sample of 7.5m tweets (500,000 users) from U.S. users to produce logistic regression models of obesity rate as a function of tweet texts.

Dashboard: To develop the dashboard, we combined data from social media and the BRFSS and imported them into an interactive display that allowed the information to be viewed in multiple formats such as timelines or maps. Timelines allow policy makers to track the response after a health program is rolled out. Color-coded maps allow them to see how obesity rate and program impact vary with location.

Results

We utilized a separately-produced user demographics prediction model to build a database aligning states with twitter users. The model excluded users with irrelevant or low confidence geotags, fake accounts, and spam campaigns. A round-robin cross-validation of the obesity-rate prediction model was performed: 50 models were built, each holding out one of the states, and the result predicted for the testing state was compared with the BRFSS reference. Pearson correlation of the resulting 50 estimates was 0.82. A large portion of the obesity rate was captured in those models. This initial result was promising and encourages us to pursue more complex models based on the tweets. This initial success provided us with the necessary data to build the PHAT dashboard.

The dashboard allows policy makers to view traditional health survey results (such as the BRFSS) with greater granularity, to query tweets while filtering through demographic variables or health program goals, and to drill down on individual messages.

Conclusions

We demonstrated that signals in social media can supplement existing survey data to provide policy makers with better tools to evaluate the success of their obesity prevention programs. The development of PHAT, especially the interactive dashboard, provides a more timely understanding of a program's impact.

Future works include tailoring PHAT to evaluate a specific program. We also plan to expand our raw data-tagging to other demographic areas (e.g. veteran status) and apply PHAT to other health behaviors (e.g. smoking).

Keywords

Social Media; Obesity; Surveillance

References

1. Park DK, Gelman A, Bafumi J. Bayesian multilevel estimation with poststratification: state-level estimates from national polls. Political Analysis. 2004;12(4):375-85.
2. Centers for Disease Control and Prevention (CDC). Behavioral Risk Factor Surveillance System Survey Data. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2011.

*Meredith Keybl
E-mail: meredith.keybl@gmail.com