A Novel Pipeline for Targeting Breast Cancer Patients on Twitter for Clinical Trial Recruitment

Jon Sligh, Hamed Abedtash, Mengye Yang, Enming Zhang, Josette Jones
Department of BioHealth, School of Informatics and Computing
Indiana University-Purdue University Indianapolis

Background and Preliminary Study

- Breast cancer is the leading form of cancer diagnosed in women which makes the discovery of an effective treatment a high priority, estimated to reach the incidence rate of 246,660 in 2016 in the US population [1].
- Scientists have developed new therapies for mitigating the disease and side effects in recent years through conducting randomized clinical trials as the gold standard clinical research method. However, recruiting individuals into clinical trials including breast cancer patients has remained a significant challenge.
- We found that clinical trials involving breast cancer patients are often terminated or not even initiated as a result of low recruitment or challenges in finding suitable qualified participants.
- Out of 525 terminated trials on breast cancer patients registered in the database, 230 (43.8%) of the terminations happened due to low or slow accrual, 34 (6.5%) due to lack of funding, and 31 (5.9%) due to toxicity concerns [3].

Objectives

- In this study, we developed and assessed a scalable framework to identify Twitter users who have breast cancer based on personal health mentions on Twitter.
- In fact, we are looking for “fingerprints” of patients’ health status on Twitter, a microblogging social networking service. This method could provide a new avenue for contacting potential study candidates for recruitment.

Methods

- First we assume that breast cancer patients who use Twitter and actively seeking for helps and support is very likely to be followers of large breast cancer communities or organizations like @Breastcancerorg, @breastcancer etc. In this way we collected twitter users who followed at least 1 of the top 40 most influential breast cancer related Twitter accounts.
- Then we retrieved their original, breast cancer related tweets for processing.
- A pipeline was developed for text processing and further analysis. The idea behind is to match cancer relevant words and phrases individually and in combinations (caner, benign, malignant, etc.), possessive terms (I, my, has, have, etc.), and supporting attributes (mass, tumor, hair loss, etc.) to determine if the user has been diagnosed with cancer.
- Using this pipeline Twitter users with high risk of having breast cancer are identified.
- The performance of the pipeline was measured by sensitivity and specificity of detecting actual breast cancer patients.

Results

- 193,024 twitter accounts were identified initially.
- 25,870,106 tweets of these accounts are collected.
- 81,429 tweets were filtered for further processing after excluding “retweets” and non-related breast cancer messages.
- 462 tweets were found using the pipeline, representing 218 unique users.
- Sensitivity and specificity of detecting actual breast cancer patients are as high as 88.7% and 91.0% with this new approach.

References