

Timothy Wells

May 18th, 2021

0890-03

Abstract:

Objective:

DUI is a tool under review and we want to compare it to a manual method which has outperformed other tools in the past. Social Media, namely Twitter presents an opportunity to gather candid data which can provide insight into substance perception and engagement.

Methods:

Retrieved and analyzed publicly available twitter messages using DUI and a variation on a pre-existing method of manual analysis.

Results:

I reviewed a total of 801 twitter messages, of which, 36% were relevant to opioid perception or experience. Of these, 84% were made by individual authors; 26% of which had to do with the author's personal experience with opioids. A majority of these messages were either substance experience descriptions or aberrant behavior with a small percentage relating to addiction. General Perceptions were split almost evenly between, messages about others' substance use, sentiments, and Info/Public Awareness.

DUI is able to consistently perform codification of terms from given text in an assumed substance abuse related context better than manual analysis but is currently not able to create context from the information derived without manual analysis of said information. To test this, I intend to perform the test suggested by Dr. Singh where context is manually recreated from DUI results using multiple tweets from one person. I assume DUI will perform well under those circumstances. DUI, although intended for reddit posts, was able to perform similar sophisticated analysis on small twitter messages.

Introduction:

Non-medical prescription opioid use is a rising public health concern illustrated by rising unintentional drug poisoning mortality rates[1]. To create a sufficient and meaningful response, there is a need to understand the changing public attitudes and beliefs about these medications.

The advent of social media represents an opportunity for candid data regarding substance abuse, with the platform “Twitter” showing notable user growth over time[2].

Drug-Use Insights (DUI) is a public and open-source web application developed/web browser application developed to perform “semantic- and exploratory analysis” on social media posts while also addressing the expanding lexicon of substance related terms[3]. I compared DUI to a manual mixed-method analysis which had exceeded similar social media analysis tools in the past[4]. My Hypothesis was that the manual method of

analysis would far exceed the ability of DUI to analyze social media text as it had done in the past.

Method:

Twitter is an online community which interacts socially via posts or messages of up to 280 characters. These posts or “Tweets” can be “liked”, replied to, or “retweeted.” Within every Tweet, the author can directly address another user via an “@”. Finally, a Tweet can categorize itself as relating to a specific topic or topics via “hashtags.” Tweepy is a Python package that allows for access to the Twitter API. For the Purpose of this project, it was used to query Twitter for relevant publicly available messages. Finally, MySQLWorkbench is a database tool for SQL development.

Within the file `get_tweets.py`, my Twitter API keys are used to provide authentication. Then I connect to the database where my tweets will be stored. Next, I have variables for the query terms and tweet date (this date is the farthest day back I want my query to go with the API limit being 7 days) and these change as needed. The next few lines perform the query over the number of tweets desired. Finally, the tweets are looped through and inserted into a table within the database.

Criteria for Tweet relevance followed the canary in the Coal Mines’ standards which were that the content of the tweet had to be in english, the content had to refer to opioid use of the search term, the context of the message had to be discernible in a meaningful manner. (I have to be able to discern if the search term within the tweet is

referring to the opioid). Retweets were filtered out instead of being stored to the database. Time range was consolidated to over two weeks. @/Screen name was also obtained.

Using a variation on the method here[4], Tweets were codified into Categories. Variations include, removal of the “Prescription opiate abuse in chronic pain patients: clinical criteria, incidence, and predictors.” article and my own individual coding refinement which resulted in the same two categories of “Personal Experience” and “General Perceptions” but each with only three sub-categories. A random selection of 2000 Tweets stored within a MySQL database were exported to a .csv file for codification. This file was manually combed through and coded in Google Sheets

From twitter using tweepy + python we grab a cross sectional sample of publicly available Twitter messages. To do this the 11 terms used here[4] plus the extra term “opioid” were queried for Tweets at a rate of 10000 results per term. Including the message itself, the date the tweet was published, and ‘@’ of the publisher were stored such that the time of the Tweet could be kept track of as well as the person Tweeting. Tweet messages were occasionally received as a concatenation of the original message and during the codification process, when I came across those, I used the Tweepy provided link to retrieve the full tweet manually.

Results:

From a total of 801 codified tweets, only 64% were found to be relevant with the remaining 287 Tweets being codified into categories. Among the Tweets codified as being Personal Experiences, a shortcoming of my method appears in Tweets which explicitly use opiates to maintain a high. This would be very precisely codified using the canary in coal mine codification but under my approach, it can only be codified as Aberrant Behavior which is not as useful if you're trying to find for example, how many individuals are using substances to get high specifically and draw a conclusion regarding that.

A majority of Personal Experience Tweets being codified as Substance experience descriptions (more than 50%) and the other majority being codified as Aberrant Behavior, suggests potential for Twitter as a platform to be a source of candid information on the experience of aberrant substance abuse. This idea would be better explored through individual Twitter profiles where aberrant behavior was identified.

That said, at times my codification of "Substance Experience Description" seems more fitting than the coalmine codification of "Side Effects of Medication" for example, tweets like "@littledes0 no not even oxycodone is working" very clearly describe an experience with a substance. An experience which would be important within the context of identifying the effectiveness of a substance vs addiction.

Very few of the Personal Experience Tweets detailed personal experience with addiction itself. This would suggest that maybe something about the Twitter format, community, or

social climate is not at this time conducive for candid conversation regarding addiction in the way some Reddit forums have shown to be in the past [5].

The portion of tweets categorized as General Perceptions were evenly split into the three sub-categories which included Commenting on others use of a substance, Public Awareness/News, and Positive and/or Negative sentiments with a negligible percent of them being codified as a pop culture reference. This seemingly reinforces the assertions made here[4] about Twitter's usefulness as a source of data on General Perceptions from individuals who made the large majority of tweets.

Both methods of codification struggled with overly simplistic yet relevant tweets. For example "Wtf I don't have an oxycodone chile", most closely fell under the codification of aberrant behavior but obviously, this does not apply to this tweet or help us categorize it in a meaningful way.

From the collection of codified tweets, users which appeared more than once were used to test DUI's ability to analyze user data over time. This was a weak area of DUI due to a lack of individual user posts from the random sample as well as the small amount of text provided for which DUI was not intended. The results returned from DUI were mostly not helpful for determining information about the user over time or comparing one user's posting to another.

That said, DUI consistently was able to extract complex categories from each Tweet and provide all possible substance abuse categories which could be derived. This exceeded the ability of the manual codification methods used by myself and[4]. This happened within a comparatively small time frame (codification taking days and DUI never taking longer than 15 minutes depending on the size of data input) with unanticipated and exceptional ease of use.

Conclusion:

Between both methods of analysis, with regards to Twitter data, context matters in terms of the data present. For example, most relevant search terms seemed to be “opioid” for its general accuracy in reference to medical substances and “fentanyl” due to George Floyd conversation and politicized Twitter sentiments. Query results from the term “Vikes” may also have been skewed by the context of sports season or Twitter being a platform for distanced massaging during the COVID-19 global pandemic where sports may normally not be as present. Manual analysis far exceeds the DUI tools in terms of time and precision needed for proper analysis. This extends to the need for exhaustive review to reduce potential biases during the categorization process.

DUI succeeds where these methods of manual analysis do with the exception of context determination given the categories DUI is able to return. If DUI is able to either consistently return categories from which context can be manually derived or determine context itself, it could potentially supplement the manual method attempted here.

Citations:

1. L. Paulozzi, Y. Xi, "Recent changes in drug poisoning mortality in the United States by urban-rural status and by drug type", *Pharmacoepidemiology and Drug Safety*, 17 (2008): 997–1005
2. P. Miller, A. Sonderlund, "Using the internet to research hidden populations", *Addiction* 105 (2010) : 1557–67, Duggan M SA. *Social Media Update 2013*. Pew Research Internet Project. 2013.
3. DUI Documentation (under review)
4. Chan B, Lopez A, Sarkar U (2015) The Canary in the Coal Mine Tweets: Social Media Reveals Public Perceptions of Non-Medical Use of Opioids. *PLOS ONE* 10(8): e0135072. <https://doi.org/10.1371/journal.pone.0135072>
5. R. Eshleman, D. Jha, and R. Singh, "Identifying Individuals Amenable to Drug Recovery Interventions through Computational Analysis of Addiction Content in Social Media", *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 839-844, 2017