Assessing the Factors that Affect the Public Engagement with Department of Transportation Twitter Accounts

Center for Accessibility and Safety for an Aging Population
Florida State University
In Partnership with Florida A&M University and University of North Florida
2525 Pottsdamer Street, Rm. 129, Tallahassee, FL 32310-6046
Phone: (850) 410-6153 Fax: (850) 410-6142
http://utc.fsu.edu/

Ayberk Kocatepe[1], Mehmet Baran Ulak[1], Javier Lores[2], Eren Erman Ozguven[1], Anil Yazici[3]

[1] Department of Civil and Environmental Engineering, Florida A&M University - Florida State University College of Engineering

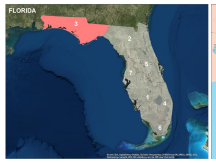
[2] Department of Computer Science, University of Central Florida [3] Department of Civil Engineering, Stony Brook University

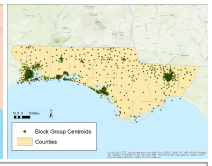
Motivation

State Departments of Transportation (DOT) utilize Twitter frequently to disseminate critical information such as accidents, road closures, and congestion to public. However, factors driving the effectiveness and reach of this information dissemination are not fully understood in the literature. Here, based on the Twitter account of Florida DOT (FDOT) District 3 region, private Twitter interaction analytics such as engagement rate and impressions, and other exogenous and endogenous variables are utilized to identify those factors influencing the success of this information dissemination. For this purpose, (a) machine learning and naïve Bayesian techniques are implemented to classify and extract information from the DOT tweets, and (b) a Tobit-based selection model is used to identify the significant endogenous (e.g. time-to-post, tweet click, tweet engagement) and exogenous factors (e.g. demographical, socioeconomic, and land use characteristics) that drive the engagement rate as an indicator of information dissemination success. This study can help transportation agencies calibrate their plans and policies towards enhancing the information dissemination, and improving the effectiveness and reach of their social media accounts.

Twitter Metrics & Available Data

- FDOT (comprising 7 districts) District 3 account has a total number of 2,602 followers and 8,679 tweets (Active since Nov. 2011).
- Twitter analytics data is only available since Sep. 2014, so , only tweets from September 2014 are utilized in this study.
- Data consists of private Twitter metrics and analytics, which can only be accessed by the account owner or manager (not available to followers)
- District 3 Twitter account provides local information: traffic accidents, roadwork disruptions, closures due to flooding, storm surges, fallen trees.





- District 3 consists of 888 block groups and 57,213 census blocks as identified by U.S Census
- Each census block includes socioeconomic data such as ethnicity, income, vehicle ownership, education level.

Twitter Analytics and their definitions

Analytics	Definition
Engagements	Total number of times someone interacted with a tweet. Clicks anywhere on the Tweet, including retweets, replies, follows, likes, links, cards, hashtags, embedded media, username, profile photo, or tweet expansion
Engagement rate	The number of engagements divided by the number of impressions
Impressions	Times people were shown a tweet in the timeline or search results
Likes	Number of people who "liked" the tweet
Retweets	Times someone retweeted the tweet
Replies	Times someone replied to the tweet
Detail expands	Clicks on the tweet to view more details
Embedded media clicks	Clicks to view a photo or video in the tweet
Follows	Times someone began following you directly from the tweet
Hashtag clicks	Clicks on hashtag(s) in the tweet
Link clicks	Clicks on a URL or card in the tweet
Permalink clicks	Clicks on the tweet's permalink (desktop only)
Shared via email	Times someone emailed the tweet to others
User profile clicks	Clicks on the name, @handle, or profile photo of the tweet's author

Objectives

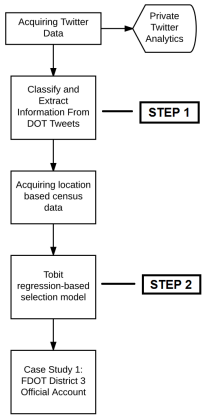
- **To understand the follower interaction with agency tweets:**

How Twitter users interact with the shared content, how many people clicked on a tweet or how many expanded a tweet, are there replies?

- **To reveal the intricacies of user engagements with tweets:**

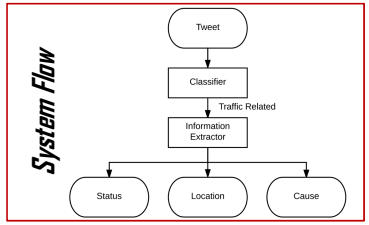
The exogenous (e.g. demographical, socioeconomic, and land use characteristics related to locations where tweets are related to) and endogenous (e.g. time-to-post, tweet click, tweet engagement) variables that influence the success of the information dissemination.

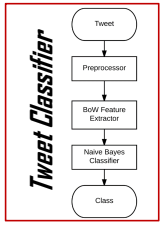
Methodology



Information Classifier and Extractor

The classifier is a machine learning (ML)-based system, which categorizes tweets based on the text as either a traffic related tweet or a non-traffic related tweet. The information extractor utilizes natural language processing (NLP) techniques to extract the status, the cause, and the street location of the detected incident. Thus, first, the flow of the system consists of sending a tweet to the classifier to determine if it is traffic related. Subsequently, if the tweet is traffic related, it is forwarded to the information extractor to determine the details about the traffic incident.

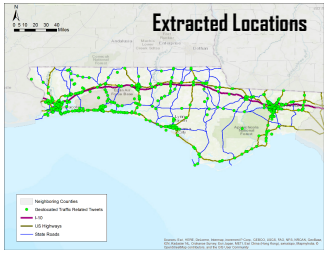


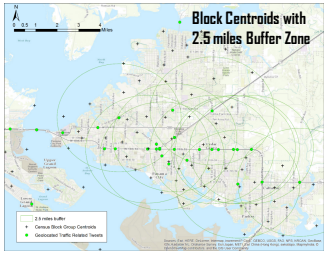


Selection Model

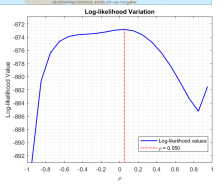
Tobit-based models can model dependent variables with censored information referring to limited/constrained observations for events. The engagement to agency twitter accounts is censored since some tweets do not have any engagements while engagements of others vary, which compels use of models account for this censoring. Moreover, natural logarithm of engagement (dependent) was used to adjust for the right-skewness (log-linear model). The selection model involves two sets of predictor variables: one to model censoring, and the other to model the occurrence frequency. That is, this model adopts unobserved stochastic censoring instead of a fixed common censoring threshold (i.e., non-stochastic) of a regular Tobit model. This approach eliminates the selection bias which begets more accurate estimates.

Results





Variables	β	SE	p	CI 95%	CI 90%	CI 80%
Selection Model						
Intercept	0.54	0.14	≈0	✓	✓	✓
AADT/1000	6.0E-05	8.0E-05	0.44	X	X	X
Below poverty %	-1.61	0.71	0.02	✓	✓	✓
College or higher education %	0.20	0.31	0.52	X	X	X
30 years old or younger %	1.02	0.55	0.06	X	✓	✓
Zero vehicle ownership %	2.75	2.82	0.33	X	X	X
Night time tweet	-0.28	0.10	0.01	✓	✓	✓
URL presence in the tweet	0.29	0.10	≈0	✓	✓	✓
Tweet content - Road work	0.14	0.11	0.17	X	X	X
PM peak time tweet	0.14	0.12	0.26	X	X	X
Regressor Model						
Intercept	0.06	0.04	0.19	X	X	✓
Impression	1.2E-04	8.0E-05	0.16	X	X	✓
Likes	0.13	0.02	≈0	✓	✓	✓
Retweet	0.35	0.02	≈0	✓	✓	✓
URL click	0.31	0.01	≈0	✓	✓	✓
Detail expansion	0.07	0.00	≈0	✓	✓	✓
User profile clicks	0.12	0.01	≈0	✓	✓	✓
Tweet content - Accident	0.24	0.06	≈0	✓	✓	✓
Tweet content - Road closure	-0.05	0.03	0.14	X	X	✓
Tweet content - Road work	-0.03	0.03	0.33	X	X	X
σ	0.39	0.03	≈0	✓	✓	✓
Number of observations	993					
Log-likelihood at zero	-1382.14					
Log-likelihood at convergence	-871.05					
Pseudo - R^2_{LL}	0.69					



- The selection equation models the probability of tweet to be engaged by the users and followers
- The regressor equation revealed the extent of each analytics' impacts on the engagement, given that tweet is engaged by users/followers.

Conclusions

- Results provide insights on **how agencies can increase engagement and activity level** by focusing on different factors, and **can promote a greater awareness** on how can make transportation-related decisions by understanding the actual engagement of people.
- **Calibrating or creating social media interaction policies** to enhance engagement performance by the findings of this study can be another important contribution from an agency perspective. For instance, **tweet posting time** can be regulated to enhance information dissemination success, and agencies might consider **standardizing the language and time** of informative tweets.
- As for **caveats**, it was discovered that there were some inconsistencies with tweet texts such as frequent typos, various abbreviations, which made it harder to detect names and extract information about the content of tweets. Google Maps' inability to handle exits and mile markers hampered the effort in identifying the locations indicated in tweets.
- As **future directions**, the 2.5 miles buffer zone used for socioeconomic and demographic variables can be extended using a sensitivity analysis regarding zone size. Furthermore, follower/user- and followers of followers-based social network analysis should provide further insights. Metropolitan cities such as New York City and Miami are considered as candidate locations for further investigation.

For information please contact: ak13y@my.fsu.edu