The Journal of Open Source Software

Sentiment Analysis of Twitter Data (SAoTD)

Evan L. Munson[1], **Christopher M. Smith**[1], **Bradley C. Boehmke**[1], **and Jason K. Freels**[1]

1 Air Force Institute of Technology

DOI: 10.21105/joss.00764

Software

- Review
- Repository
- Archive

Submitted: 17 April 2018
Published: 04 June 2018

SAoTD is an R interface to the Twitter API and can be used to acquire tweets based on user selected #hashtags. The package will clean and tidy the Twitter data, determine the latent topics within the tweets utilizing Latent Dirichlet Allocation (LDA), determine a sentiment score using the Bing lexicon dictionary and output visualizations.

The package is available on GitHub and archived on Zenodo. The package was developed with the intention of the user creating a personal Twitter account which will then allow a user to access the twitter API throught the Twitter Developers Account site. Once a user has access to the Twitter Developers Account they will have the ability to create an application which will then provide the user with access tokens. These access tokens will then allow the user to begin acquiring tweets usint eh SAoTD package.

The package is laid out in five different categories: Acquire, Explore, Topic Analysis, Sentiment Calculation, and Visualizations.

- Acquire allows a user to acquire tweets of their choosing by accessing the Twitter API.
- Explore provides functions to tidy, explore unigrams, bigrams, tri-grams, in addition to bigram netwrorks and correlation networks.
- Topic analysis allows a user to explore the latent topics buried within the tweets.
- Sentiment calculation utilizes the Bing lexicon dictionary to score the text (Hu and Liu 2004).
- Visualizations allow the user to better understand the twitter sentiment.

The package utilizes tidy dataferames and therefore depends on the tidyverse package (Wickham 2017) and additionally uses the tidytext package (Silge and Robinson 2017). The number of latent topics is determined using the ldatuning package (Nikita 2016) and the latent dirichlet allocation (LDA) topics is determined using the topicmodels package (Grün and Hornik 2011).

The SAoTD package has research applications in many disciplines which intend to use twitter text and sentiment analysis. The package was created to quickly determine the sentiment of twitter and to inform analysts on the opinions contained within tweets.

References

Grün, Bettina, and Kurt Hornik. 2011. "topicmodels: An R Package for Fitting Topic Models." *Journal of Statistical Software* 40 (13):1–30. https://doi.org/10.18637/jss.v040.i13.

Hu, Minqing, and Bing Liu. 2004. "Mining and summarizing customer reviews." In *Proceedings of the Tenth Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 168–77. ACM.

Nikita, Murzintcev. 2016. *Ldatuning: Tuning of the Latent Dirichlet Allocation Models Parameters*. https://CRAN.R-project.org/package=ldatuning.

Silge, Julia, and David Robinson. 2017. *Text Mining with R a Tidy Approach*. Edited by Nicole Tache. 1st ed. Sebastopol: O'Reilly. http://tidytextmining.com/.

Wickham, Hadley. 2017. *Tidyverse: Easily Install and Load the 'Tidyverse'*. https://CRAN.R-project.org/package=tidyverse.