

Sentiment Analysis of Twitter Data (saotd)

Evan L. Munson¹, Christopher M. Smith¹, Bradley C. Boehmke¹, and Jason K. Freels¹

¹ Air Force Institute of Technology

DOI: [10.21105/joss.00764](https://doi.org/10.21105/joss.00764)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted: 17 April 2018

Published: 27 February 2019

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

`saotd` is an R package that provides a programmatic interface to the Twitter API and can be used to acquire tweets based on user-specified #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 create output visualizations.

The package is available on [GitHub](#) and archived on [Zenodo](#). To configure the package a user must follow these steps:

- Creating a personal [Twitter](#) account if they don't already have one.
- Enable a [Twitter Developers Account](#).
- Create an application which will provide them with API access tokens required by the `saotd` package to function.

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 networks 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 & Liu, 2004).
- Visualizations allow the user to better understand the sentiment of the tweets.

The package utilizes tidy dataframes and therefore depends on the `tidyverse` package (Wickham, 2017) and additionally uses the `tidytext` package (Silge & 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 & Hornik, 2011).

The `saotd` package has research applications in many disciplines which need to access tweets from the Twitter platform and carry out sentiment analyses. This package was created to quickly determine the sentiment of Twitter and to inform analysts on the opinions contained within tweets.

References

Grün, B., & Hornik, K. (2011). `topicmodels`: An R package for fitting topic models. *Journal of Statistical Software*, 40(13), 1–30. doi:[10.18637/jss.v040.i13](https://doi.org/10.18637/jss.v040.i13)

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth acm sigkdd international conference on knowledge discovery and data mining* (pp. 168–177). ACM.

Nikita, M. (2016). *Ldatuning: Tuning of the latent dirichlet allocation models parameters*. Retrieved from <https://CRAN.R-project.org/package=ldatuning>

Silge, J., & Robinson, D. (2017). *Text Mining with R a Tidy Approach*. (N. Tache, Ed.) (1st ed., p. 178). Sebastopol: O'Reilly. Retrieved from <http://tidytextmining.com/>

Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'*. Retrieved from <https://CRAN.R-project.org/package=tidyverse>