

shankarz.TextST package

Natarajan Shankar

2015-03-18

Package Description

- Package: shankarz.TextST
- Type: Package
- Title: A Twitter Recommendation Engine
- Version: 1.0
- Date: 2015-03-18
- Author: Natarajan Shankar
- Maintainer: Natarajan Shankar <shankarz@stanford.edu>
- Description: shankarz.TextST : A tweet recommendation engine
- Depends: R (>= 2.10)
- License: GPL (>= 2)
- LazyData: true
- Imports: RCurl, ROAuth, plyr, slam, tm, twitteR, stringr
- Suggests: testthat, knitr
- VignetteBuilder: knitr

shankarz.TextST - Abstract

The **shankarz.TextST** package implements an information retrieval and processing system based upon Twitter tweets. In social engineering, data exchanged between people carries information that can then be broken down and sorted. Twitter is one such source of inter-people information. Sorted data from Twitter can then be used to help make smart, correlative future decisions. The use of past Twitter data to make future projections represents the crux of this project.

Given a collective set of tweets that correspond to a specific search string in a specific geography in a specific time frame:

shankarz.TextST uses the TF-IDF algorithm to prioritize the ASCII tags contained in the collective set and forms a prioritized dataset. 2). The prioritization of the ASCII tags relies upon the well known TF-IDF algorithm (**Reference:** <http://en.wikipedia.org/wiki/Tf\T1\textendashidf>) 3). Given a new search string and the pre-existing prioritized ASCII tags, shankarz.TextST searches the prioritized dataset for relevant tweets.

The utility of this package is that a user with a general idea of what they are looking for in Twitter can use a very general search string, feed it as a filter into a large prioritized dataset, and then be able to retrieve a list of the top N relevant tweets. This project, conducted for as part of the requirements for the Stats290 curriculum, is a rudimentary foray into the area of filtered search. *While the prioritized datasets may not exist today, it is expected that data summarization and filtered search will become an important aspect in future years.*

Visible functions available in shankarz.TextST package

1. **** twitterAuthenticate()** Set up and complete Twitter authentication ******
**** Description**** twitterAuthenticate() Authenticates user session with Twitter

Usage `twitterAuthenticate()` **Details** Prerequisite task (to be completed at dev.twitter.com) The dev.twitter.com application control panel provides the ability to generate an OAuth access token for the owner of this application. Visit the -My applications- page by navigating to appsdev.twitter.com, or hover over the profile image in the top right hand corner of the twitter site and select -My applications-.

The Applications page contains a list of the applications, along with a button to create a application. Select the application you wish to generate a token for, and click on its title. Click on the -Create my access token- button, and an authorized access token and secret will be generated for this account and the current application.

Once the access tokens have been generated, they are to be into this function. Knowing the the `shankarz.TextST` application will subsequently make requests on behalf of this R session.

Once the prerequisite task is complete, `twitterAuthenticate` will succeed in accessing the Twitter API with the pre-generated access token and secret:

Author Natarajan Shankar

**** References**** <http://dev.twitter.com>

**** Example function call**** `twitterAuthenticate()`

2. ****** `twitterBuildDatabase()` Get tweets from Twitter, given user credentials and tokens******

Description

`twitterBuildDatabase()`

Gets tweets from within an area within miles of a location specified by latitude and longitude

Usage

```
twitterBuildDatabase(search.string = "", lat = "", long = "", miles = "", since = "", numberOfTweets = 1)
```

Arguments

1. `search.string` An optional search string to query Twitter.
2. `lat` A string of latitude numerics.
3. `long` A string of longitude numeric.
4. `miles` The number of miles for radius of search (as string).
5. `since` The start date for the Twitter search ("yyyy-mm-dd").
6. `numberOfTweets` Number of tweets requested from Twitter.

Details

This function conducts a search that pulls tweets that originated within specified miles of a location and were tweeted after the date specified by the `since` parameter. Optionally, this function also accepts a search string.

Value

A data.frame of twitter tweets.

Example function call `message("To run twitterBuildDatabase, it is mandatory to have have run TwitterAuthenticate()") twitterAuthenticate()`

```
twitterBuildDatabase(search.string = "Giants", lat="37.794108", long="-122.39511",  
                      miles="15", since="2014-10-01", numberOfTweets = 100)
```

““

1. `# generateTweetRecommendations()` Given a new tweet, return n recommended tweets.**

Description

`generateTweetRecommendations()`

- Recommends tweets in the corpus given a seed tweet.

Usage

`generateTweetRecommendations(tweetDF, tweet = NULL, num_recs = 10)`

Arguments

1. `tweetDF`

- A data.frame returned from the prior step that called `twitterBuildDatabase()`.

2. `tweet`

- The seed string for which recommendations will be returned for.

3. `num_recs`

- The number of recommendations requested.

Details

This function constructs a DocumentTermMatrix (reference: package `tm`) that is weighted by TF-IDF in order to show similar words to any given tweet. Given a seed tweet, this function adds the tweet to the corpus, calculates TF-IDF on the entire set, and recommends n similar tweets based on text similarity.

Value

tweets recommended for the seed tweet.

Example function call

`generateTweetRecommendations(sfTweets, tweet="sourdough", num_recs=10)`

Author Natarajan Shankar

** References** <http://dev.twitter.com>

Running the entire code sequence in 3 easy function calls

```
# Load the needed library
library(shankarz.TexST)
```

```
twitterAuthenticate()
```

```
df <- twitterBuildDatabase("sf", lat="37.7941",
                           long="-122.395",
                           miles="15",
                           since="2015-01-01",
                           numberOfTweets=500)
```

```
seedTweet <- "food at the wharf"
generateTweetRecommendations(sfTweets, seedTweet)
```

```
## [1] "Sometimes Fishermans Wharf reminds me of Crete. #fishingboats #greece #crete #sanfrancisco #sf"
## [2] "Today #sfmarinfoodbank @ SF-Marin Food Bank https://t.co/ln3BIPDkS5"
## [3] "Today @skoutapp and @sfmfoodbank #volunteer @ SF-Marin Food Bank https://t.co/L7xySQHTU1"
## [4] "#volunteer #volunteering #sffoodbank #sanfranciscofoodbank #oranges @ SF-Marin Food Bank https://t.co/ln3BIPDkS5"
## [5] "Look at our HUGE @RocketLawyer group skipping orientation bc we're repeat volunteers at the SF"
## [6] "@rocketlawyer playing it cool before volunteering at @sfmfoodbank #toocoolforschool @ SF-Marin Food Bank"
## [7] "Newest food trend in SF: teriyaki @XFINITY (spotted also by @serifluous & @jordanstaniscia)"
## [8] "First time in SF (@ San Francisco International Airport (SFO) - @flysfo in San Francisco, CA) I"
## [9] "Wouldnt be a bad view from the desk.. #SF #sanfran @ Google San Francisco https://t.co/IbxU0b7y"
## [10] "These city folk sho love their rooftop experiences http://t.co/I9ZEQ7FJU1 #sf #bayarea http://t.co/I9ZEQ7FJU1"
```

```
seedTweet <- "downtown homeless"
generateTweetRecommendations(sfTweets, seedTweet)
```

```
## [1] "Why do I not have LTE in downtown SF @VerizonSupport"
## [2] "walked around downtown sf and it was really cool"
## [3] "I wish I ate more before coming here. We are being educated about what to do when we become homeless"
## [4] "Downtown's 800 J Lofts sell to real estate investment firm http://t.co/tWY0c5q501"
## [5] "Just passed a homeless man digging through garbage streaming SFPD on his phone. Only in SF."
## [6] "@Hoos1492 I am an enthusiastic promoter of \"reactivation villages\" for homeless neighbors in SF"
## [7] "A site created (GoFundMe-like fundraising pages) for homeless people. #PrivatizePoverty #SFTown"
## [8] "Rally for #Tibet downtown SF yesterday. @ Union Square, San Francisco https://t.co/6UBb4ybV7c"
## [9] "#karl has come to downtown #fog #cold #sf @ San Francisco, California https://t.co/AXtYUJPZD1"
## [10] "Work lobby. #SF #WorkingInSF #MySecondDay #Thursday #Work #Downtown @ San Francisco, California"
```